Credit Assignment: Challenges and Opportunities in Developing Human-like Learning Agents

Thuy Ngoc Nguyen[1], **Chase McDonald**[2], **Cleotilde Gonzalez**[2]
[1]Department of Computer Science, University of Dayton
[2] Department of Social and Decision Sciences, Carnegie Mellon University
ngoc.nguyen@udayton.edu, chasemcd@cmu.edu, coty@cmu.edu

Abstract

Temporal credit assignment is the process of distributing delayed outcomes to each action in a sequence, which is essential for learning to adapt and make decisions in dynamic environments. While computational methods in reinforcement learning, such as temporal difference (TD), have shown success in tackling this issue, it remains unclear whether these mechanisms accurately reflect how humans handle feedback delays. Furthermore, cognitive science research has not fully explored the credit assignment problem in humans and cognitive models. Our study uses a cognitive model based on Instance-Based Learning Theory (IBLT) to investigate various credit assignment mechanisms, including equal credit, exponential credit, and TD credit, using the IBL decision mechanism in a goal-seeking navigation task with feedback delays and varying levels of decision complexity. We compare the performance and process measures of the different models with human decision-making in two experiments. Our findings indicate that the human learning process cannot be fully explained by any of the mechanisms. We also observe that decision complexity affects human behavior, but not model behavior. By examining the similarities and differences between human and model behavior, we summarize the challenges and opportunities for developing learning agents that emulate human decisions in dynamic environments.

Introduction

Learning the relationship between actions and outcomes is essential for behavioral adaptation and decision-making in dynamic environments (Gonzalez, Lerch, and Lebiere 2003). A difficult learning challenge arises when a decision maker must make a series of decisions without receiving feedback until the end of the sequence. The value of each decision can only be determined retrospectively after learning the final outcome of the task, and this learning process is important for improving future decisions. This problem, known as the *temporal credit assignment*, determines how credit should be assigned to intermediate actions within a sequence (Minsky 1961). The gap between when a decision is made and when the outcome of such a decision is observed is known as “feedback delay” in dynamic decision-making research (Brehmer 1989), and it is one of the most challenging problems in learning to improve decisions over time in dynamic situations (Gonzalez, Lerch, and Lebiere 2003; Gonzalez, Fakhari, and Busemeyer 2017).

Computational sciences have proposed a number of approaches to handle delayed feedback. One of the most prominent mechanisms for addressing the credit assignment problem is the temporal difference (TD) from the reinforcement learning (RL) literature (Sutton 1985; Sutton and Barto 2018). TD approaches enable an agent to predict the value of intermediate states in the absence of final feedback and use prediction errors over small intervals to update their future predictions. Many RL algorithms use TD methods (Van Seijen et al. 2009; Hasselt 2010; Xu, van Hasselt, and Silver 2018), including several state-of-the-art deep RL algorithms (Mnih et al. 2015; Van Hasselt, Guez, and Silver 2016; Hessel et al. 2018).

Recent advances in deep RL algorithms have allowed artificial intelligence (AI) agents to reach a level of human performance that has not been possible before in a variety of complex decision-making tasks (Wong et al. 2021). However, these RL agents appear to be less adaptable to novel situations compared to humans, who can quickly learn many different tasks and quickly generalize knowledge from one task to another (Pouncy, Tsividis, and Gershman 2021). Although previous studies have shown that RL models can account for human behavior in some dynamic decision tasks (Simon and Daw 2011; Gershman and Daw 2017), none of the current models can account for this human ability to adapt rapidly in situations with delayed feedback, and AI agents are often inadequate to explain and predict adaptation and learning in complex environments as humans do (Lake et al. 2017; Pouncy, Tsividis, and Gershman 2021). Therefore, concerns have been raised that the advancement in RL algorithms is mainly focused on solving computational problems efficiently and optimally rather than replicating the way humans actually learn (Botvinick et al. 2019; Lake et al. 2017). The purpose of this paper is to highlight the challenges and insights in the development of human-like learning agents that adapt and learn in dynamic decision-making situations with delayed feedback. This research uses RL agents and cognitive models of decision-making based on IBLT (Gonzalez, Lerch, and Lebiere 2003) and analyses of human actions in the same tasks.

Background

Research in AI has a longstanding goal of replicating various human behaviors in computational form so that the machine's behavior would be indistinguishable from that of a human (Lake et al. 2017; Turing 1950). Building accurate replications of human decisions, a "cognitive clone" of a human cognitive decision process, is essential to anticipate human error and to create personalized and dynamic digital assistants, as has been recently shown in various applications of cognitive models (Somers, Oltramari, and Lebiere 2020; Gonzalez et al. 2021). Little effort has been dedicated to investigating how to build "human-like" models that consider the cognitive plausibility and diversity of human behavior (Gonzalez 2023). Consequently, a significant challenge for AI research is to develop systems that can replicate human learning behavior (Lake et al. 2017).

Given that cognitive architectures have been developed to represent an integrated view of the cognitive capacities of the human mind (Anderson et al. 2004; Anderson and Lebiere 2014), previous research has explored how well models align with humans in tasks involving feedback delays (Walsh and Anderson 2011, 2014). In particular, TD credit assignment methods have been incorporated into cognitive architectures to emulate how humans process feedback delays in sequential decision-making tasks (Fu and Anderson 2006). Other cognitive modeling research suggests that people evaluate intermediate states in terms of future rewards, as predicted by TD learning (Walsh and Anderson 2011). However, these studies have primarily focused on the similarities between neural processes and computational mechanisms, leaving room for further investigation and comparison of observed human behavior and credit assignment mechanisms in sequential decision-making tasks.

Computational cognitive models, which are based on cognitive architectures, have demonstrated the ability to represent human decision-making processes in a variety of tasks. In particular, Instance-Based Learning (IBL) models that rely on the theoretical principles of IBLT (Gonzalez, Lerch, and Lebiere 2003) have been used to emulate human binary choices (Gonzalez and Dutt 2011) and decisions in more complex dynamic resource allocation tasks such as the Internet of Things (Somers, Oltramari, and Lebiere 2020), cybersecurity (Gonzalez et al. 2020), multistate gridworld tasks (Nguyen and Gonzalez 2020, 2021), and multi-agent settings are required to build real-time interactivity between models and humans (Nguyen, Phan, and Gonzalez 2023a). IBLT provides a single general algorithm and mathematical formulation for memory retrieval that is based on the well-known ACT-R cognitive architecture (Anderson and Lebiere 2014). It has emerged as a comprehensive theory of the cognitive process by which humans make decisions based on experience in dynamic environments (Gonzalez 2023; Gonzalez and Dutt 2011; Hertwig 2015; Nguyen, Phan, and Gonzalez 2023b). In IBLT, the question of temporal credit assignment is addressed through a feedback process, but the development and comparison of particular mechanisms for credit assignment that mimic human behavior in IBLT are still in the early stages of exploration (Gonzalez 2023; Nguyen and Gonzalez 2020).

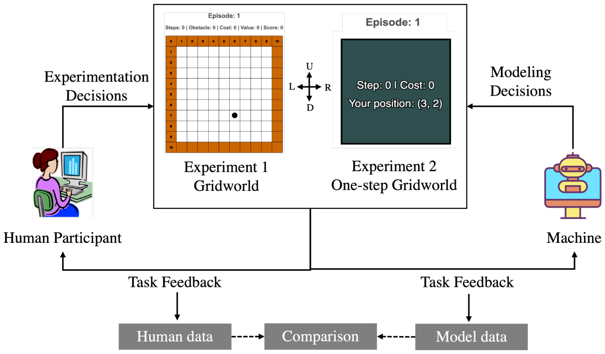

Figure 1: Experimental scenarios.

Goals and Research Approach

In this study, we investigate the problem of credit assignment in a gridworld task that involves delayed feedback at varying levels of decision complexity. To achieve this, we use an IBL model that implements different credit assignment mechanisms, including IBL-Equal (equal credit), IBL-Exponential (exponential credit), and IBL-TD (Nguyen, Phan, and Gonzalez 2023a) (TD credit) using the IBL decision mechanism, as well as Q-learning, the fundamental TD algorithm in RL (Watkins and Dayan 1992; Sutton and Barto 2018). We analyze the predictions from these models on human performance under the same gridworld tasks. We adopt a commonly used approach previously employed in cognitive research (Busemeyer and Diederich 2010; Gonzalez 2017) to validate cognitive models with human data. The process is illustrated in Figure 1.

Our goal is to gain an in-depth understanding of the ability of these models to produce *human-like* behavior and to expose challenges and insights into the development of human-like AI agents. To develop insights into the various methods of temporal credit assignment, we consider three credit assignment mechanisms implemented in the IBL model: equal credit, exponential credit, and TD credit using the IBL decision mechanism. We compare results obtained from model simulations and human experiments to determine how closely the models represent human behavior. For this study, we conducted two human experiments that provided different visual representations of the same gridworld tasks. The analysis of performance and optimal actions helps determine which credit assignment mechanisms produce behavior similar to that of humans and which result in more optimal and effective actions than those of humans.

After comparing the results of these models with the outcomes and process measures of human decisions in the two experiments, we have concluded that an IBL-Equal model that equally assigns credit to all decisions is able to match human performance more closely than other models, including IBL-Exponential, IBL-TD and Q-learning. Initially, the learning speed of IBL-TD and Q-learning models was inferior to that of humans, but eventually, the models surpassed human performance.

We then examine the differences between the strategies employed by humans and those used by models. This allows us to understand the nuances of human behavior, what is and is not captured by AI agents, and the human characteristics that may hinder optimal decisions. Our findings show that humans have the ability to create a mental model of a task based on its visual representation. In addition, humans tend to approach tasks more strategically than models do. Our experiments reveal that humans have concepts and strategies that influence their behavior, which are not captured by models. We also found that humans spend less time exploring environments than models, which leads to long-term suboptimality compared to TD models.

Taken together, this work brings us closer to understanding the general algorithms of credit assignment that can be used to generate human-like models. It also highlights the challenges that researchers need to address to capture the initial strategic behavior that humans might carry from one task to another, as well as potential strategies to enhance human decisions in sequential decision-making tasks with delayed feedback. The results and insights from this research will need to be used in future work to develop human-like learning agents and to utilize these models to support human activities in future AI systems.

References

Anderson, J. R.; Bothell, D.; Byrne, M. D.; Douglass, S.; Lebiere, C.; and Qin, Y. 2004. An integrated theory of the mind. *Psychological review*, 111(4): 1036.

Anderson, J. R.; and Lebiere, C. J. 2014. *The atomic components of thought*. Psychology Press.

Botvinick, M.; Ritter, S.; Wang, J. X.; Kurth-Nelson, Z.; Blundell, C.; and Hassabis, D. 2019. Reinforcement learning, fast and slow. *Trends in cognitive sciences*.

Brehmer, B. 1989. Feedback delays and control in complex dynamic systems. In *Computer-based management of complex systems*, 189–196. Springer.

Busemeyer, J. R.; and Diederich, A. 2010. *Cognitive modeling*. Sage.

Fu, W.-T.; and Anderson, J. R. 2006. From recurrent choice to skill learning: A reinforcement-learning model. *Journal of experimental psychology: General*, 135(2): 184.

Gershman, S. J.; and Daw, N. D. 2017. Reinforcement learning and episodic memory in humans and animals: an integrative framework. *Annual review of psychology*, 68: 101–128.

Gonzalez, C. 2017. Decision-making: a cognitive science perspective. *The Oxford handbook of cognitive science*, 1: 1–27.

Gonzalez, C. 2023. Building Human-Like Artificial Agents: A General Cognitive Algorithm for Emulating Human Decision-Making in Dynamic Environments. *Perspectives on Psychological Science*, 17456916231196766.

Gonzalez, C.; Aggarwal, P.; Cranford, E. A.; and Lebiere, C. 2021. Adaptive Cyberdefense with Deception: A Human-AI Cognitive Approach.

Gonzalez, C.; Aggarwal, P.; Lebiere, C.; and Cranford, E. 2020. Design of dynamic and personalized deception: A research framework and new insights.

Gonzalez, C.; and Dutt, V. 2011. Instance-based learning: Integrating decisions from experience in sampling and repeated choice paradigms. *Psychological Review*, 118(4): 523–51.

Gonzalez, C.; Fakhari, P.; and Busemeyer, J. 2017. Dynamic decision making: Learning processes and new research directions. *Human factors*, 59(5): 713–721.

Gonzalez, C.; Lerch, J. F.; and Lebiere, C. 2003. Instance-based learning in dynamic decision making. *Cognitive Science*, 27(4): 591–635.

Hasselt, H. V. 2010. Double Q-learning. In *NeurIPS*, 2613–2621.

Hertwig, R. 2015. Decisions from experience. *The Wiley Blackwell handbook of judgment and decision making*, 1: 240–267.

Hessel, M.; Modayil, J.; Van Hasselt, H.; Schaul, T.; Ostrovski, G.; Dabney, W.; Horgan, D.; Piot, B.; Azar, M.; and Silver, D. 2018. Rainbow: Combining improvements in deep reinforcement learning. In *AAAI*, volume 32.

Lake, B. M.; Ullman, T. D.; Tenenbaum, J. B.; and Gershman, S. J. 2017. Building machines that learn and think like people. *Behavioral and brain sciences*, 40.

Minsky, M. 1961. Steps toward artificial intelligence. *Proceedings of the IRE*, 49(1): 8–30.

Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A. A.; Veness, J.; Bellemare, M. G.; Graves, A.; Riedmiller, M.; Fidjeland, A. K.; Ostrovski, G.; et al. 2015. Human-level control through deep reinforcement learning. *nature*, 518(7540): 529–533.

Nguyen, T. N.; and Gonzalez, C. 2020. Effects of Decision Complexity in Goal-seeking Gridworlds: A Comparison of Instance-Based Learning and Reinforcement Learning Agents. In *Proceedings of the 18th intl. conf. on cognitive modelling*.

Nguyen, T. N.; and Gonzalez, C. 2021. Theory of Mind From Observation in Cognitive Models and Humans. *Topics in Cognitive Science*.

Nguyen, T. N.; Phan, D. N.; and Gonzalez, C. 2023a. Learning in Cooperative Multiagent Systems Using Cognitive and Machine Models. *ACM Transactions on Autonomous and Adaptive Systems*, 18(4): 1–22.

Nguyen, T. N.; Phan, D. N.; and Gonzalez, C. 2023b. SpeedyIBL: A comprehensive, precise, and fast implementation of instance-based learning theory. *Behavior Research Methods*, 55(4): 1734–1757.

Pouncy, T.; Tsividis, P.; and Gershman, S. J. 2021. What Is the Model in Model-Based Planning? *Cognitive Science*, 45(1): e12928.

Simon, D. A.; and Daw, N. D. 2011. Environmental statistics and the trade-off between model-based and TD learning in humans. In *NeurIPS*, 127–135.

Somers, S.; Oltramari, A.; and Lebiere, C. 2020. Cognitive Twin: A Cognitive Approach to Personalized Assistants. In *AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering (1)*.

Sutton, R. S. 1985. Temporal Credit Assignment in Reinforcement Learning.

Sutton, R. S.; and Barto, A. G. 2018. *Reinforcement learning: An introduction*. MIT press.

Turing, A. 1950. Mind. *Mind*, 59(236): 433–460.

Van Hasselt, H.; Guez, A.; and Silver, D. 2016. Deep reinforcement learning with double q-learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30.

Van Seijen, H.; Van Hasselt, H.; Whiteson, S.; and Wiering, M. 2009. A theoretical and empirical analysis of Expected Sarsa. In *2009 ieee symposium on adaptive dynamic programming and reinforcement learning*, 177–184. IEEE.

Walsh, M. M.; and Anderson, J. R. 2011. Learning from delayed feedback: neural responses in temporal credit assignment. *Cognitive, Affective, & Behavioral Neuroscience*, 11(2): 131–143.

Walsh, M. M.; and Anderson, J. R. 2014. Navigating complex decision spaces: Problems and paradigms in sequential choice. *Psychological bulletin*, 140(2): 466.

Watkins, C. J.; and Dayan, P. 1992. Q-learning. *Machine learning*, 8(3-4): 279–292.

Wong, A.; Bäck, T.; Kononova, A. V.; and Plaat, A. 2021. Multiagent deep reinforcement learning: Challenges and directions towards human-like approaches. *arXiv preprint arXiv:2106.15691*.

Xu, Z.; van Hasselt, H. P.; and Silver, D. 2018. Metagradient reinforcement learning. *NeurIPS*, 31: 2396–2407.