Hierarchical Reinforcement Learning for Decision Support in Health Care

Caroline Strickland^{†,*}, Daniel Lizotte[†]

[†] University of Western Ontario

Keywords: Machine Learning, Hierarchical Reinforcement Learning, Sequential Decision Support

1. Research Problem

Organizations often collect data to support their operational decisions. Many of these organizations (like health care) make decisions sequentially but operate hierarchically. For example, an administrator may make a decision about staffing that affects decisions across an organization while individual care givers within that organization make decisions impacting individual patients. Furthermore, the objectives of decision-making at different levels may be different, even though the decisions affect each other. Many health-care organizations collect data about both types of decisions, for example in Electronic Medical Record (EMR) databases.

Reinforcement learning (RL) is a category of machine learning focused on agents taking actions to maximize the notion of cumulative reward based on feedback from the environment [1]. RL has been successfully applied to many sequential decision-making problems, especially within health care [2], however current applications of RL to health care data do not acknowledge or leverage the benefits of its hierarchical structure. Therefore, they cannot reveal how decision-making at different levels interact, and they cannot accommodate unique reward signals at different levels of decision-making. Thus, they cannot be used to learn optimal decisions.

By exploiting information about how layers affect each other, we can learn to make better decisions within each layer. Hierarchical reinforcement learning (HRL) [3] is a powerful approach containing multiple frameworks for solving long-horizon problems with sparse and delayed rewards. At its core, HRL decomposes a RL problem into a hierarchy of subtasks to be solved individually using RL. While the tiered-task structure that HRL employs fits our problem well, existing frameworks are not quite suitable for two primary reasons. First, in traditional HRL problems, subtasks are largely disconnected from each other; the decisions made in one subtask will not affect other subtasks within the problem. Since we want to capitalize on the inter-connectedness of layers, we must consider the decisions made in a certain layer and use them to make better decisions within other layers of the hierarchy. Second, in traditional HRL problems, all subtasks operate within the same state-space. Because we will be dealing with multiple layers that make different types of decisions, many layers will have their own disconnected state space. This introduces new challenges in both complexity reduction and model formalization.

Currently, there is no RL methodology for hierarchical decision-support of this kind. This work will propose RL methods for hierarchical decision support and apply these methods to health care data from long-term care facilities. While we will be using health care as our testbed, this hierarchical decision-support model could potentially be employed within many other hierarchical organizations.

*cstrick4@uwo.ca

2. Formalizing a HRL Framework for Sequential Decision Support

HRL develops principled ways of temporal and state abstraction to reduce the dimensionality for sequential decision-making [4]. The primary idea behind HRL is the notion of an “abstract action”. These abstract actions, unlike the primitive actions of RL that directly affect the environment, modify the internal state of the agent to activate different behavioral sub-policies. There are a number of frameworks used within HRL, however we will be using the basic structure of the MAXQ framework developed by Dietterich [5] as a foundation for our work. The MAXQ framework can be summarized as decomposing the value function of a Markov Decision Process (MDP) into combinations of value functions of smaller MDPs, a finite set of subtasks where each individual subtask must be formalized as a termination predicate, a pseudo-reward, and a set of actions. MAXQ’s advantage over other frameworks is that has recursive optimality (the policy for a parent task is optimal given the learned policies of its children).

While the foundational concepts of MAXQ are beneficial to us, we must still develop a novel HRL framework to address the complex nature of our problem. In doing this, we must consider the inability of current frameworks to deal with batch learning, multi-agent coordination, and modeling concurrent activities. Formalizing our framework hinges on incorporating solutions for each of these problems.

It has been assumed in the formalization of classic HRL frameworks that the agent can interact with the environment while learning. In health care settings, it is incredibly expensive to collect data, and significant missteps could cause damage to people or hardware. Thus, online learning is essentially forbidden within this context. Currently there is little prior work in developing batch learning algorithms that allow hierarchical semi-Markov decision processes (SMDPs), which generalize MDPs by allowing the state transitions to occur in continuous irregular times, to be trained from an existing dataset collected from a stochastic behavior policy in a similar way to how it would be collected within a health care domain [6]. Fortunately, there has been research into hierarchical Q-value iteration (HQI) by Zhao et al. [4] showing that it is possible to learn optimal policies for any valid hierarchical decomposition given a fixed dataset collected from a flat stochastic behavioral policy. We plan to incorporate the fundamentals behind HQI to allow for batch learning.

Our hierarchy is multi-layered with each layer containing at least one decision-making agent, and thus is a multi-agent system. Research in Hierarchical multi-agent reinforcement learning (HMARL) is limited, however we plan on incorporating the work of Ghavamzadeh et al. [7] and their introduction of “Cooperative HRL”. Cooperative HRL is an extension of the MAXQ framework to the multi-agent case. In this framework, agents are cooperative and homogenous. In Cooperative HRL the learning is decentralized, with each agent learning to coordinate to achieve the same overarching goal. We plan on extending Cooperative HRL to take into account the distinct state-spaces of layers for reduced complexity.

Finally, agents must learn to operate concurrently on different layers within separate state-spaces (concurrent learning). To handle this issue, we plan to incorporate ideas from Marthi et al. [8] and their “Concurrent ALisp”, a language for HRL in multi-effector problems. Concurrent ALisp allows multi-threaded partial programs, where each thread corresponds to a task. These threads can be created or destroyed over time, and prior knowledge about coordination between threads can be included in the partial program. If not, the threads coordinate their choices to maximize the global utility function.

While there are several other issues to address due to the complexity of our problem (a major one being the allowance of flexibility in the policies), these three make up a significant portion of the functionality within our novel HRL framework. Our framework would aid not only health care, but also similarly-structured organizations and would offer insight not yet accessible in existing HRL frameworks.

References

- [1] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- [2] C. Yu, J. Liu, and S. Nemati. “Reinforcement learning in healthcare: A survey”. In: *arXiv preprint arXiv:1908.08796* (2019).
- [3] A. G. Barto and S. Mahadevan. “Recent advances in hierarchical reinforcement learning”. In: *Discrete event dynamic systems* 13.1 (2003), pp. 41–77.
- [4] T. Zhao and M. Gowayyed. “Algorithms for batch hierarchical reinforcement learning”. In: *arXiv preprint arXiv:1603.08869* (2016).
- [5] T. G. Dietterich. “An overview of MAXQ hierarchical reinforcement learning”. In: *International Symposium on Abstraction, Reformulation, and Approximation*. Springer. 2000, pp. 26–44.
- [6] M. K. Bloch. “Reducing commitment to tasks with off-policy hierarchical reinforcement learning”. In: *arXiv preprint arXiv:1104.5059* (2011).
- [7] M. Ghavamzadeh, S. Mahadevan, and R. Makar. “Hierarchical multi-agent reinforcement learning”. In: *Autonomous Agents and Multi-Agent Systems* 13.2 (2006), pp. 197–229.
- [8] B. Marthi, S. J. Russell, D. Latham, and C. Guestrin. “Concurrent hierarchical reinforcement learning.” In: *IJCAI*. 2005, pp. 779–785.