Curriculum Design for Machine Learners in Sequential Decision Tasks

(Extended Abstract)

Bei Peng
Washington State University
Pullman, WA, USA
beipeng.peng@gmail.com

James MacGlashan
Brown University
Providence, RI, USA
jmacglashan@gmail.com

Robert Loftin
North Carolina State University
Raleigh, NC, USA
rtloftin@ncsu.edu

Michael L. Littman
Brown University
Providence, RI, USA
mlittman@cs.brown.edu

David L. Roberts
North Carolina State University
Raleigh, NC, USA
robertsd@csc.ncsu.edu

Matthew E. Taylor
Washington State University
Pullman, WA, USA
taylorm@eecs.wsu.edu

ABSTRACT

Existing machine-learning work has shown that algorithms can benefit from curricula—learning first on simple examples before moving to more difficult examples. While most existing work on curriculum learning focuses on developing automatic methods to iteratively select training examples with increasing difficulty tailored to the current ability of the learner, relatively little attention has been paid to the ways in which *humans* design curricula. We argue that a better understanding of the human-designed curricula could give us insights into the development of new machine-learning algorithms and interfaces that can better accommodate machine- or human-created curricula. Our work addresses this emerging and vital area empirically, taking an important step to characterize the nature of human-designed curricula relative to the space of possible curricula and the performance benefits that may (or may not) occur.

Keywords

Curriculum Design; Curriculum Learning; Sequential Decision Tasks; Human-Agent Interaction

1. INTRODUCTION

Humans acquire knowledge efficiently through a highly organized education system, starting from simple concepts, and then gradually generalizing to more complex ones using previously learned information. Similar ideas are exploited in animal training [10]—animals can learn much better through progressive task shaping. Recent work [1, 5, 6] has shown that machine-learning algorithms can benefit from a similar training strategy, called *curriculum learning*. Rather than considering all training examples at once, the training data can be introduced in a meaningful order based on their apparent simplicity to the learner, such that the learner can build up a more complex model step by step. The agent will be able to learn faster on more difficult examples after it has mastered simpler examples. This training strategy was shown to drastically affect learning speed and generalization in supervised learning settings [5, 6].

While most existing work on curriculum learning (in the context of machine learning) focuses on developing automatic methods to iteratively select training examples with increasing difficulty tailored to the current ability of the learner, how *humans* design curricula is one neglected topic. Taylor et al. [15] first showed that curricula work in reinforcement learning (RL) domains via transfer learning by gradually increasing the complexity of tasks. Narvekar et al. [8] developed different methods to automatically generate novel source tasks for a curriculum and showed that such curricula could be successfully used for transfer learning in multiagent RL domains. Svetlik et al. [14] proposed to use reward shaping [9] to automatically construct effective curricula given a set of source tasks. However, none of their work investigates human-designed curricula. We believe non-expert users may be able to design successful curricula by considering which examples are "too easy" or "too hard," similar to how humans are taught with the *zone of proximal development* [17]. A better understanding of the curriculum-design strategies used by humans may help us design machine-learning algorithms and interfaces that better accommodate natural tendencies of human trainers.

Another motivation for this work is the increasing need for non-expert humans to teach autonomous agents new skills without programming, given that more robots and virtual agents become deployed. Published work in Interactive Reinforcement Learning [2, 3, 4, 7, 11, 12, 16] has shown that reinforcement learning (RL) [13] agents can successfully speed up learning using human feedback, demonstrating the significant role humans play in teaching an agent to learn a (near-) optimal policy. Curriculum design is another paradigm that people could use to teach the agent to speed up learning. In this paradigm, the human teacher needs to design a sequence of source tasks for the agent to train on, rather than directly giving evaluative feedback to the agent. Decades of research in human education have emphasized the role of curriculum design to promote learning. Therefore, we argue that human-designed curricula are a critical area for the field of human-agent interaction. Specif-

Appears in: *Proc. of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), S. Das, E. Durfee, K. Larson, M. Winikoff (eds.), May 8–12, 2017, São Paulo, Brazil.*

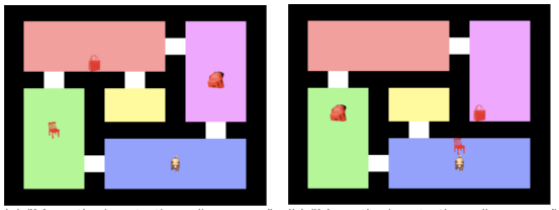

Figure 1: The (a) target environment #1 and (b) target environment #2 and their corresponding commands used in our study.

ically, this work focuses on understanding non-expert human teachers rather than finding the most efficient way to solve our sequential decision problem—future work will investigate how to adapt machine-learning algorithms to better take advantage of this type of non-expert guidance. We believe this work is the first to explore how non-expert humans approach designing curricula in the context of sequential decision tasks.

In this work, we introduce and define the curriculum design problem in the context of sequential decision tasks. In our sequential decision domain, an agent must learn tasks in a simulated home environment. The tasks are specified via text commands and the agent is trained with reinforcement and punishment. The goal of a curriculum is to allow an agent to improve learning.

Existing work [8] has shown that a multistage curriculum can speed up learning when the final (*target*) task is too difficult for the agent to learn from scratch, we aim to explore the effect of curricula when the target task is not too hard to directly learn. We hypothesize that more benefits of curricula could be found as the complexity of the target task increases. To explore how non-experts generate curricula, we task non-expert humans with designing a curriculum for an agent and evaluate the curricula they produce.

2. OUR DOMAIN

Our domain is a simplified simulated home environment of the kind shown in Figure 1. The domain consists of four object classes: agent, room, object, and door. The agent can deterministically move one unit in the four cardinal directions and pushes objects by moving into them. The objects are chairs, bags, backpacks, or baskets. Rooms and objects can be red, yellow, green, blue, and purple. Doors (shown in white in Figure 1) connect two rooms so that the agent can move from one room to another. The possible commands given to the agent include moving to a room (*e.g.*, "move to the red room") and taking a specified object to a room (*e.g.*, "move the red bag to the yellow room"). The agent learns to follow these text commands via an automated trainer's reinforcement and punishment feedback.

3. CURRICULUM DESIGN

In curriculum learning, the goal is to generate a sequence of n tasks, $M_1, M_2, \ldots, M_n$, for an agent to train on. The agent should train on these n tasks and then train on the pre-defined target task, M_t. The curriculum is successful if learning on the target task M_t is faster with the curriculum than without it. A more difficult goal is to construct a sequence such that training on the entire $n+1$ tasks is faster than training directly on the final task, M_t. In our setting, speed is measured via the number of trainer feedbacks required to learn.

In this work, a set of source tasks (94) is provided to be slected into a curriculum.[1] Each task M_i is defined by 1) a training environment with an initial state and 2) a text command. To study the effect of the target task's complexity on the performance of curricula, we designed two target task room layouts with the same command as shown in Figure 1. The second target task is harder than the first one because there are more competing hypotheses on the agent's way to the goal state in the second target task.

4. SIMULATION RESULTS

We generated four sets of random curricula of lengths $n = \{1, 2, 3, 4\}$. There were 200 curricula for each of the four sets. Each curriculum was generated by randomly selecting a sequence of tasks from the provided 94 source tasks. Each of these 800 curricula was evaluated 20 times and compared to directly learning the target task. One main result is that compared to directly learning each of the two target tasks, all four sets of random curricula could reduce the amount of feedback required to learn. Feedback required could be reduced more in the second, harder target task than in the first, demonstrating that more benefits of curricula could be found as the target task's complexity increases.

5. HUMAN SUBJECTS RESULTS

To study whether non-expert humans can design good curricula for an agent, we developed an empirical study in which participants were asked to design a set of training assignments for the dog to help it quickly learn to complete the final target assignment (the harder one was chosen). We considered data from 80 unique workers on Amazon Mechanical Turk.

One main result is that compared to directly learning the target task, less feedback was required for the agent to 1) master the intended task, and 2) learn all tasks within the curricula (including the target task) after training on curricula designed by participants.[2] Thus, the more difficult goal of curriculum design was achieved. It is also worth noting that participants were not given any feedback on the quality of the curricula they created, which demonstrates that non-expert humans can successfully design curricula that result in better overall agent performance than learning from scratch, even in the absence of relative curricula evaluation.

We also find that non-expert users can discover and follow salient principles when selecting tasks in a curriculum. Specifically, they prefer 1) isolating complexity, 2) selecting the simplest environments they can to introduce one complexity at a time, 3) choosing environments that are most similar to the target environment, and 4) introducing complexity by building on previous tasks rather than backtracking to introduce a new type of complexity. These principles can be highly useful for the design of new machine-learning algorithms that accommodate human teaching strategies.

[1] Asking humans or agents to *construct* source tasks is an interesting problem left for future work.

[2] In these experiments, participants only design the curricula — an automated trainer provides explicit feedback on 50% of the agent's actions.

REFERENCES

[1] Y. Bengio, J. Louradour, R. Collobert, and J. Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48, 2009.

[2] S. Griffith, K. Subramanian, J. Scholz, C. Isbell, and A. L. Thomaz. Policy shaping: Integrating human feedback with reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 2625–2633, 2013.

[3] W. B. Knox and P. Stone. Interactively shaping agents via human reinforcement: The TAMER framework. In *The Fifth International Conference on Knowledge Capture*, September 2009.

[4] W. B. Knox and P. Stone. Reinforcement learning from simultaneous human and MDP reward. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, pages 475–482, 2012.

[5] M. P. Kumar, B. Packer, and D. Koller. Self-paced learning for latent variable models. In *Advances in Neural Information Processing Systems*, pages 1189–1197, 2010.

[6] Y. J. Lee and K. Grauman. Learning the easy things first: Self-paced visual category discovery. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1721–1728, 2011.

[7] R. Loftin, B. Peng, J. MacGlashan, M. L. Littman, M. E. Taylor, J. Huang, and D. L. Roberts. Learning behaviors via human-delivered discrete feedback: modeling implicit feedback strategies to speed up learning. *Autonomous Agents and Multi-Agent Systems*, 30(1):30–59, 2015.

[8] S. Narvekar, J. Sinapov, M. Leonetti, and P. Stone. Source task creation for curriculum learning. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems*, Singapore, May 2016.

[9] A. Y. Ng, D. Harada, and S. Russell. Policy invariance under reward transformations: Theory and application to reward shaping. 1999.

[10] B. F. Skinner. Reinforcement today. *American Psychologist*, 13(3):94, 1958.

[11] H. B. Suay and S. Chernova. Effect of human guidance and state space size on interactive reinforcement learning. In *2011 Ro-Man*, pages 1–6, 2011.

[12] K. Subramanian, C. L. Isbell Jr, and A. L. Thomaz. Exploration from demonstration for interactive reinforcement learning. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems*, pages 447–456, 2016.

[13] R. S. Sutton and A. G. Barto. *Introduction to reinforcement learning*, volume 135. MIT Press Cambridge, 1998.

[14] M. Svetlik, M. Leonetti, J. Sinapov, R. Shah, N. Walker, and P. Stone. Automatic curriculum graph generation for reinforcement learning agents. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 2016.

[15] M. E. Taylor, P. Stone, and Y. Liu. Transfer Learning via Inter-Task Mappings for Temporal Difference Learning. *Journal of Machine Learning Research*, 8(1):2125–2167, 2007.

[16] A. L. Thomaz and C. Breazeal. Reinforcement learning with human teachers: Evidence of feedback and guidance with implications for learning performance. In *AAAI*, pages 1000–1005, 2006.

[17] L. S. Vygotsky. *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press, 1978.