How Do Humans Teach: On Curriculum Design for Machine Learners

(Doctoral Consortium)

Bei Peng
School of Electrical Engineering and Computer Science
Washington State University
Pullman, Washington, USA
bpeng@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. This thesis aims to better understand the curriculum-design strategies followed by non-experts when teaching the agent, and leverage the findings to develop new machine-learning algorithms and interfaces that better accommodate natural tendencies of human trainers. We discuss completed work on this topic, including the definition of a curriculum-design problem in the context of sequential decision tasks, analysis of how different curricula affect agent learning in a Sokoban-like domain, and results of a user study that explores whether non-experts generate such curricula. Finally, we also present directions for future work.

Keywords

Curriculum Design; Curriculum Learning; Sequential Decision Tasks; Learning from Reinforcement; Human-Agent Interaction; Crowdsourcing Experiments

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 [8]—animals can learn much better through progressive task shaping. Recent work [1, 3] 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.

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. 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. Published work in Interactive Reinforcement Learning [2, 4] (IRL) has shown that reinforcement learning (RL) [9] 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. As more robots and virtual agents become deployed, the majority of teachers will be non-experts.

Taylor [10] first proposed that curricula should be automatically designed in an RL context, and that we should try to leverage human knowledge to design more efficient curricula. As far as we know, there is very limited work on exploring how non-expert humans approach designing curricula in the context of sequential decision tasks. Therefore, the goal of this thesis work is to 1) explore how different curricula affect agent learning in different sequential decision-making domains, 2) better understand non-expert human teachers in designing curricula, and 3) develop new machine-learning algorithms that better take advantage of this type of non-expert guidance.

2. COMPLETED WORK

Before putting efforts into exploring human teaching strategies when designing curricula in our Sokoban-like test domain [5], we investigated how humans train virtual agents in the same domain using reward and punishment feedback and how to design a better representation of the agent to speed up learning [7]. We developed an adaptive speed agent that was able to adapt its action execution speed to learn more efficiently from human feedback. It could encourage more explicit feedback from human trainers in areas of the state space where the agent had more uncertainty about how to

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.*

act. Our user-study results showed that the adaptive speed agent dominated different fixed speed agents on several measures of performance. We presented this work at *AAMAS-16* [7] and a short video summarizing this work is available at `https://www.youtube.com/watch?v=AJQSGD_XPrk`.

Curriculum design is another paradigm people could use to teach the agent to speed up learning. Our initial work [6] introduced a curriculum-design problem in the context of sequential decision tasks and conducted a user study to explicitly explore how non-expert humans go about assembling curricula. The goal of our curriculum design problem is to design a sequence of source tasks $M_1, M_2, \dots, M_n$ for an agent to train on such that it can complete the given target task M_t quickly with little explicit feedback. Each source task M_i was defined by a training environment, initial state, and a command to complete in that environment. We started with providing subjects a library of environments with different level of complexities to select. Participants could select environments and corresponding commands in any order to design their own curricula. The results of our empirical study showed that non-expert users could 1) successfully design curricula that result in better overall agent performance (relative to learning from scratch), even in the absence of feedback on their quality; and 2) discover and follow salient patters when selecting tasks in a curriculum—an attribute we plan to leverage in the design of machine learning algorithms in the future.

In our extended abstract to be presented at *AAMAS-17*, we present new analysis of how different curricula affect agent learning in the same Sokoban-like domain. 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 environments and corresponding commands from the provided library of environments, allowing repeats. Each of these 800 curricula was evaluated 20 times and compared to directly learning the target task. The simulation results showed that 1) different curricula can have substantial impacts on training speeds, 2) longer curricula may result in better agent performance in learning all tasks within the curricula (including the target task), 3) more benefits of curricula can be found as the target task's complexity increases, and 4) the method for providing reward feedback to the agent as it learns within a curriculum does not change which curricula are best.

3. FUTURE WORK

For future work, we can speculate on ways of generalizing our findings to more complex task domains: 1) choose a reward feedback strategy that minimizes the number of actions needed for the agent to complete the more complex task (e.g., robot navigation tasks), where the training time is an important performance metric, 2) incorporate the salient principles (e.g., isolating complexity) we found about humans when designing curricula into the automatic process of generating useful source tasks, 3) improve the interface to guide the non-experts to design better curricula, and 4) build new machine learning algorithms with inductive biases that favor the types of curricular changes that human trainers tend to use.

Another key question to address is exploring how good are people at understanding what an agent can do when designing curricula. We hypothesized that people who interpret what the agent currently knows correctly could design better curricula. Thus, we plan to design a new user study where we show users an agent performing some tasks and ask them some questions regarding what the agent learned to do thereafter. We can then study whether people's understanding of what an agent knows affect their task selection in curriculum design, which is related to the zone of proximal development [11] for curriculum design.

Acknowledgements

This research has taken place in part at the Intelligent Robot Learning Lab, Washington State University, supported in part by JCATI (WSU003346), NASA (NNX16CD07C), NSF (IIS-1149917 and IIS-1643614), and NVidia.

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] 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-Volume 1*, pages 475–482, 2012.

[3] M. P. Kumar, B. Packer, and D. Koller. Self-paced learning for latent variable models. In *NIPS*, pages 1189–1197, 2010.

[4] 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.

[5] J. MacGlashan, M. L. Littman, R. Loftin, B. Peng, D. L. Roberts, and M. E. Taylor. Training an agent to ground commands with reward and punishment. In *Proceedings of the AAAI Machine Learning for Interactive Systems Workshop*, 2014.

[6] B. Peng, J. MacGlashan, R. Loftin, M. L. Littman, D. L. Roberts, and M. E. Taylor. An empirical study of non-expert curriculum design for machine learners. In *Proceedings of the Interactive Machine Learning workshop (at IJCAI)*, July 2016.

[7] B. Peng, J. MacGlashan, R. Loftin, M. L. Littman, D. L. Roberts, and M. E. Taylor. A need for speed: Adapting agent action speed to improve task learning from non-expert humans. In *AAMAS*, 2016.

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

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

[10] M. E. Taylor. Assisting transfer-enabled machine learning algorithms: Leveraging human knowledge for curriculum design. In *The AAAI 2009 Spring Symposium on Agents that Learn from Human Teachers*, March 2009.

[11] L. S. Vygotsky. *Mind in society: The development of higher psychological processes*. Harvard university press, 1980.