Training Agents by Crowds

Elnaz Nouri
Computer Science Department, University of Southern California
Institute for Creative Technologies
Playa Vista, California 90094

Abstract

On-line learning algorithms are particularly suitable for developing interactive computational agents. These algorithm can be used to teach the agents the abilities needed for engaging in social interactions with humans. If humans are used as teachers in the context of on-line learning algorithms a serious challenge arises: their lack of commitment and availability during the required extensive training. In this work we address this challenge by showing how "crowds of human workers" rather than "single users" can be recruited as teachers for training each learning agent. This paper proposes a framework for training agents by the crowds. The focus of this proposal is narrowed by using Reinforcement Learning as the human guidance method for teaching agents how to engage in simple negotiation games (such as the Ultimatum Bargaining Game and the Dictator Game).

Motivation

The main idea behind this research proposal is to use crowds of people to allow seamless consistent and extensive amount of training for agents to learn their optimal policy. This on-line learning paradigm is most suitable when the agents are learning to engage in interactions with humans, for example, in the context of human-computer dialog conversations such as negotiation or bargaining. The approach introduced here uses multiple humans as trainers and addresses the challenges with previous approaches that used single users(teachers) for training computer agents. Human guidance techniques that use single users for training learning agents (mostly robots), often face challenges when they are put in practice because human teachers often do not commit to the extensive ((and boring) amount of training needed for the algorithm to converge to the optimal policy. The single user human teacher approaches would also have issues because it is not possible for the agent to explore all possible actions during their interaction with the user. As a result they might not be able to learn the optimal policy needed for the task. I became interested to work on these problems because I encountered with them in my own work towards developing culture sensitive negotiating agents (NT11; NGT12; NGT14). My thesis work so far has aimed at developing culture sensitive policies for interactive agents that can engage with humans from diverse cultural background. My focus was on their decision making behavior. Previously I tried to do this with two different methods:

1. Learning from off-line human-human interaction data (NT11; NGT14)
2. Learning from the people from people from the cultures (referred to as experts in (AN04)'s method for apprenticeship learning) interactively using inverse reinforcement learning. (NGT12)

The first approach can be criticized because it is not a human-agent interaction and people's behavior when interacting with one another might not be similar to how they would act when interacting with a computer agent. The second approach however required extensive training iterations because it was not possible to have real humans interact with our system for training, we had to build Simulated Users (GHL) based on the behavioral data (RPOFZ91) we had in order to be able to perform the training. The proposed framework in this paper can address the mentioned problems through crowd scouring. That's the main concern in my work but along the way, I am also interested to study how the diversity of the individual workers (teachers) affects the training of the agents and interactions with the system. The issues discussed here are relevant to problems in multiple areas of work, including dialog systems, human-robot interaction (HRI), mixed-initiative assisting agents, and task-or domain centric versions of the assisting agents.

Background and Related work

Human guidance techniques have previously been combined with reinforcement learning (RL) (SB98) in the form of demonstration (AN04) or reward (KS09). Although reinforcement learning is a suitable paradigm for guiding machines with human expertise and help, many challenges arise when it's put to practice. The RL training can be done off-line or on-line (interactively). In the off-line approach; the RL requires a lot of data for training and converging to an optimal policy. Theoretically one could train a policy for performing a certain task by using a very large corpus but such large corpora usually are not available and even if they do exist they might not be properly annotated or prepared for training an agent. For most real world applications, it is often the case that even if a suitable dataset is available most

likely it would not include all possible policies and user behaviors. For the on-line approach, real users have been previously shown to help RL systems train their policy (GJT$^+$) but the fact that the human user as the instructor for the system in these scenarios needs to commit to training period, represents an effort-time overload on behalf of the human user. Many users have been shown to quit the task during the training process (Pae). The training process is slow and a lot of interaction between the users and RL agents are needed. Exploration of the RL agent also becomes an issue when user is interacting with real users. If the agents explore too much then the system would be very frustrating to use for humans. On the other hand, if the system doesn't explore enough then it wouldn't learn the optimal policy. One solution to the mentioned problems is to use Simulate Users (GHL) which is an algorithm that uses user features and previous behavior to generate behavior in interaction with the systems. The main criticism in using Simulated Users is that often building a simulated user is just as hard as building a policy. Another problem is not knowing what kinds of simulated users would help build the best policies. Even when Simulated Users are made and used to learn policies there's always a need to have real users evaluate the learned policies.

Research Questions

Can a crowd of users be used to train a RL agent? Can this approach of using multiple teachers for training agents address the shortcomings of the previously discussed approaches (e.g. simulated users)? These are the main questions addressed in this proposal. In the context of my work the goal is to investigate the possibility of effectively training Reinforcement Learning agent with multiple teachers as they find their optimal policy to play simple negotiation and bargaining games. Reinforcement Learning algorithms require extensive amount of iterations before converging to the optimal policy and the games chose are tasks in which the social concerns have been shown to play a determining role in the way people behave.

The other main question that I look into is what the best arrangement (pooling or queuing strategy) is for setting up multiple teachers that might take turns voluntarily or involuntarily when they interact with computer agents. If the teachers are to take turn one after another they need to be provided some context for resuming the training task. Investigation is needed for determining how much context is needed to preserve be efficiency and consistency of the training process at the same time.

Another inevitable but interesting follow up question is on how diverse crowds of workers can be brought to train agents. This is interesting because when multiple teachers are used it is very likely that they might not all be similar and whether or not this can affect training should be investigated. In the context of my work for this proposal this means investigation of different patterns of behavior in games which can be influence by many factors such as personal and cultural traits.

Proposed Research

A framework needs to be set up for training agents by using a crowd of teachers. So far I have set up a web application through which agents can play the games with humans. I have also considered designing a mobile application because it enables extensive interaction with the users. I am still investigating whether the scattered interactions through the mobile application would have the same quality of a short term committed interaction. Currently the teachers are assigned one after another but I am investigating this approach and comparing it with other methods of pooling the teachers. Questions on how to provide the context of the interaction to the newly assigned teachers should also be investigated. In order to evaluate the framework at its current stage, two sets of comparisons are needed. One should compare the performance of this framework with other approaches. In the context of this work the comparison would be between this framework and systems that have been trained off-line or with Simulated Users for training negotiating agents. The second set of comparisons would compare the performance of single teacher agents with that of multiple agents.

Proposed Experiments

I suggest that we train agents by using the crowd to play few simple games with humans. By using the decision making games (The Dictator Game and the Ultimatum Game in multiple rounds (Cam03)) that I have previously used in my work (NGT12), I can compare the performance of the new framework to the previous work that I have done with single teachers.

Challenges

Based on my previous experiments, I think the following would be the major challenges in this work:

Recruitment of Workers

Concurrency As the technology and theoretical work on crowd sourcing evolves many issues with concurrently recruiting workers are being addressed, however, in our previous experiment this aspect has been a challenge and we need to resolve it in order to carry out the proposed experiments.

Diversity Considering the focus of my thesis on capturing culture specific behavior (a source of diversity in people and perhaps the teachers of the agents), I have made several efforts for recruiting people from different countries on Amazon Turk. Although we had no problem recruiting people from US and India on Amazon Turk, this was not the case for other cultures even when we left the task accessible for extended period of time, we were not able to recruit people from many countries in Europe and Africa or the Middle East. A potential solution seems to be using different crowd sourcing websites.

References

Pieter Abbeel and Andrew Y Ng. Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the*

twenty-first international conference on Machine learning, page 1. ACM, 2004.

Colin Camerer. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press, 2003.

Kallirroi Georgila, James Henderson, and Oliver Lemon. User simulation for spoken dialogue systems: learning and evaluation. Citeseer.

Milica Gašic, Filip Jurcıcek, Blaise Thomson, Kai Yu, and Steve Young. On-line policy optimisation of spoken dialogue systems via live interaction with human subjects.

W Bradley Knox and Peter Stone. Interactively shaping agents via human reinforcement: The tamer framework. In *Proceedings of the fifth international conference on Knowledge capture*, pages 9–16. ACM, 2009.

Elnaz Nouri, Kallirroi Georgila, and David Traum. A cultural decision-making model for negotiation based on inverse reinforcement learning. In *Proceedings of the 34th annual meeting of the Cognitive Science Society*, pages 2097–2102. Annual Conference of the Cognitive Science Society, 2012.

Elnaz Nouri, Kallirroi Georgila, and David Traum. Culture-specific models of negotiation for virtual characters: multi-attribute decision-making based on culture-specific values. *Journal of AI and Society*, 1(1):87–111, 2014.

Elnaz Nouri and David Traum. A cultural decision-making model for virtual agents playing negotiation games. In *Proceedings of the International Workshop on Culturally Motivated Virtual Characters*. 11th International Conference on Intelligent Virtual Agents, 2011.

Tim Paek. Reinforcement learning for spoken dialogue systems: Comparing strengths and weaknesses for practical deployment.

Alvin E Roth, Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir. Bargaining and market behavior in jerusalem, ljubljana, pittsburgh, and tokyo: An experimental study. *The American Economic Review*, pages 1068–1095, 1991.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*, volume 1. Cambridge Univ Press, 1998.