Greedy Agents and Interfering Humans – An artwork making humans meddle with a life in the machine

Tatsuo Unemi[1], Daniel Bisig[2] and Philippe Kocher[2]

[1]Soka University, Hachioji, Japan, unemi@soka.ac.jp
[2]Zurich University of the Arts, Switzerland

Abstract

This article introduces an experimental artwork that employs a reinforcement learning algorithm as core element for an interactive and aesthetic experience. The learning algorithm involves a simple navigation task for a single agent. The agent's learning process is made perceivable to visitors by animating and visualizing a massive particle system on which the agent's memory acts as force field. Through interaction, visitors can either facilitate or hamper the agent's learning process. The goal of the artwork is to convey in a playful manner the increasingly intertwined coexistence between humans and artificially intelligent entities.

Introduction

Greedy Agents and Interfering Humans is an experimental artwork that aims to convey through interactivity and visualization the increasingly intertwined coexistence between humans and artificially intelligent entities. The artwork takes the form of a tabletop installation that invites visitors to interact with an agent that learns to navigate a simulated environment. The installation renders the learning process perceivable through visualization and exposes it to manipulation through interaction. As a result, the learning process unfolds before the visitors and becomes part of the aesthetic expression of the artwork.

The artwork employs reinforcement learning (RL) as learning method for the agent. RL is a paradigm of animal learning which has been researched for more than one hundred years in the fields of psychology and ethology. It serves as a powerful framework to explain how an individual's behavior is modified through experience. At its core, RL is based on a trial and error approach combined with a reward scheme. During learning, behaviors that lead to positive experiences are reinforced whereas behaviors with negative consequences are suppressed.

In late of 19th century, Thorndike (1898) initiated scientific research into these phenomena in the context of social psychology. Several decades later, Skinner (1953) conducted systematic experiments on pigeons and rats following a *Behaviorism* approach. In these experiments, the animals changed their behavior to increase positive experiences (being fed) and avoid negative experiences (electric shocks).

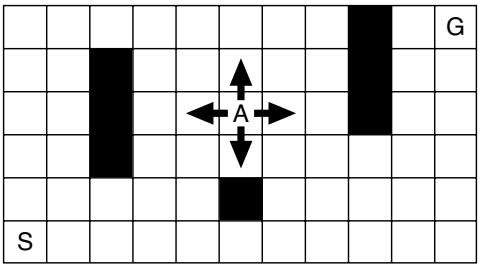

Figure 1: Navigation in a grid world as learning task. Outlined rectangles represent traversable space and filled rectangles represent obstacles. The labels are: S for Start, G for Goal, and A for Agent.

In 1980s, with the advent of powerful computational resources, it become feasible to adopt principles of reinforcement learning in the context of machine learning. From the early 1990s on, substantial research on computational forms of reinforcement learning has been conducted by Sutton and Barto (2018).

Implementation

The artwork employs a form of reinforcement learning that is based on the classic framework of Q-Learning (Watkins, 1989). Following this framework, a single agent navigates a small grid-world in which each cell either represents an empty space or an obstacle. In the current version, the grid world is two-dimensional and 11×6 cells in size (see Fig. 1). The agent can chose among four discrete actions (up, down, left and right) to move from one grid cell to the next. Movement is only possible into an empty grid cell. The agent's task is to move from a start cell to a target cell with a minimum number of actions. Based on findings from an evolutionary optimization of learning parameters (Unemi et al., 1994), the agent's probability of choosing a random action (exploration) instead of an optimal action (exploitation) is gradually reduced as learning progresses. A learning

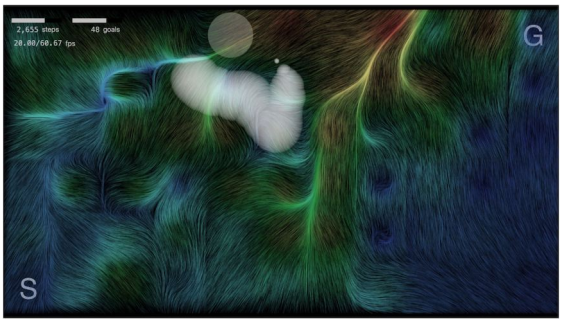

Figure 2: A still image of the visualization. A semitransparent white circle depicts the position of the agent. The white worm-like shape depicts the trace of visitor's hand. *Start* and *Goal* cells are indicated by the letters S and G, respectively.

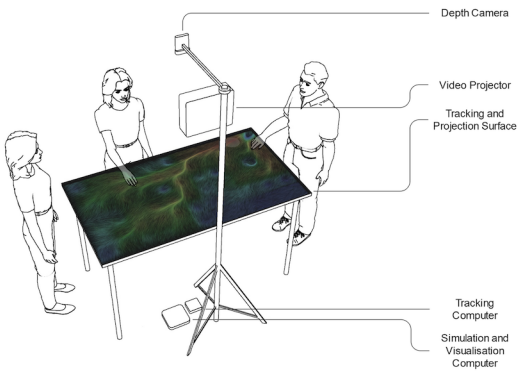

Figure 3: Tabeltop installation setup.

episode begins with the agent placed on a *Start* cell and ends when the agent has either exceeded a maximum number of actions or has reached a *Goal* cell. The agent receives a positive reward when it reaches the *Goal* cell. The agent's performance is measured by how often it reaches the *Goal* cell. As a result, the agent eventually find the shortest path from *Start* to *Goal* cell.

Learning progresses with the agent memorizing the values to each of its actions at each of the grid cell locations. The higher the value of an action at a cell position, the more likely the agent will take this action. When the agent reaches the *Goal* cell, the value of the last action taken is propagated backward from the *Goal* cell to previous positions along the agent's path. To accelerate this propagation, a replay mechanism is employed that causes the agent to randomly recall previous navigation steps from a memory pool. This mechanism is similar to the *Dyna* architecture proposed by Sutton (1990).

The objective of our project is an artistic application of reinforcement learning rather than an improvement of its learning performance within an engineering context. For this reason, we decided to work with a learning algorithm that has low computational demands and is easy to implement. While the original Q-learning algorithm fulfills these two criteria, it was deemed unsuitable for our application because its training progresses too slowly for visitors to witness in realtime. The integration of a rehearsal mechanism increases the speed of training and thereby rectifies this issue. Recent progress in machine learning has led to the emergence of much more powerful mechanisms than the one we are employing in this work. One example of such mechanisms is *Deep Q-Networks* (Mnih et al., 2015). These networks can adapt to complex environments and complex learning tasks such as the ones tackled by *AlphaGo* (Silver et al., 2016).

The agent's memory is represented as a vector field in which the direction and length of each vector is derived from the action values at the corresponding cell location. The vector is calculated by summing the orthogonal directions of the four discrete actions scaled by their respective action value. The visualization of the agent's memory is implemented by means of a particle system in which each particle moves across the grid world. The motions of the particles result from the forces the vector field excerpts on them. Several hundred thousand particles are rendered as short line segment with a color that depends on the particles speed. In the installation, the particle animation is projected on a table surface. Figure 2 shows a still image of a particle animation.

Interaction is based on tracking the visitors' hand positions on top of the table surface. A schematic depiction of several visitors interacting with the installation is shown in Figure 3. The visitors' hands are detected by means of a distance camera mounted above the table and pointing vertically down. The positions are derived from the front-most points of the hands' contours and mapped to the cells of the grid world with which the positions overlap. Through interaction, visitors can cause various effects that either hinder or accelerate the agent's learning process. These effects include: temporarily guiding or blocking the agent's navigation, creating additional obstacles or removing existing obstacles, tracing navigation paths that alter the agent's memory.

Outlook

For future versions of the installation, it is planned to complement the visualization of reinforcement learning with a sonification. The properties of the cell properties (e.g. it's action values, state as obstacle or empty space) are taken as control parameters for driving a real-time sound synthesis algorithm. Unlike visualization, which reflects the current state of the agent's learning, sonification will serve the purpose of revealing the history of the learning process. By placing their hands on specific grid cells, visitors could for example listen to how some of the cell's properties changed over the course of multiple agent navigations.

References

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533.

Silver, D., Huang, A., Maddison, C., Guez, A., Sifre, L., Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D. (2016). Mastering the game of go with deep neural networks and tree search. *Nature*, 529:484–489.

Skinner, B. F. (1953). *Science and human behavior*. The Macmillan Company, New York.

Sutton, R. S. (1990). Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In *7th International Conference on Machine Learning*, pages 216–224. Morgan Kaufmann.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press, 2nd edition.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, 2(4):i–109.

Unemi, T., Nagayoshi, M., Hirayama, N., Nade, T., Yano, K., and Masujima, Y. (1994). Evolutionary differentiation of learning abilities-a case study on optimizing parameter values in q-learning by a genetic algorithm. In *Artificial life IV*, pages 331–336. MIT Press.

Watkins, C. J. C. H. (1989). *Learning from delayed rewards*. PhD thesis, University of Cambridge.