Reinforcement Learning: A Broad Look

Christopher Cabrera
Florida State University
Cc12d@my.fsu.edu

Abstract

Reinforcement learning is the term used for agents that learns from a series of reinforcements of awards and punishments (Russell and Norvig). Reinforcement learning is a quickly growing facet of the artificial intelligence field. This paper will introduce the reader to the methods and applications of reinforcement learning agents.

Introduction

When children are learning to interact with their environment, they do so by a series of trial and error. They are awarded by the correct moves and punished by the incorrect moves. Our understanding of the world we lives in comes from testing it, interpreting the results of that test, and using the knowledge we've gained to make proper decisions in future endeavors. That is the premise that fuels reinforcement learning. By utilizing this method of learning, reinforced learning agents can very closely mimic how humans handle problems. Reinforcement learning is used in artificial intelligence by awarding an agent for certain "good" states and penalizing the agent for other "bad" states. As an agent continually tests its states and finds the combination of states that consistently return the highest rewards, an agent will make logical decisions. The more the agent explores its environment and tests other ways to increase reward; it will in turn learn better methods to reaching its goal state.

Background

Reinforcement learning has been used to solve a number of artificial intelligence problems; including robot navigation, network administration, automated surveillance, and much more (Hardesty). The first known use of a reinforced-learning agent was in 1959. This is also the same algorithm that first used machine learning in a checkers program written by Arthur Samuel (Russell and Norvig). Since then, there have been a number of advancements in reinforcement learning that has made it a very powerful tool in the artificial intelligence community. There are a number of algorithms that have been used in reinforcement learning, among them are the Q-Learning method and the SARSA method.

Q-Learning is a powerful reinforcement learning algorithm that works by estimating the values of state-action pairs (Watkins). The Q-learning algorithm is guaranteed to converge to the correct Q-values in a stable environment.

SARSA is similar to Q-Learning except that the max reward in the next state is not always used for updating the Q-values.

Unlike supervised learning algorithms, these agents can take a basic blank slate with little to no information about their environment and learn to not only succeed, but also excel at a task that they have learned themselves with simple positive and negative enforcement.

Algorithm

As an example of a reinforcement learning algorithm, I will explain a Q-Learning method of a search algorithm. The search algorithm is a simple grid with solid walls on the outside given a reward of -1 and a goal state randomly chosen somewhere in the grid with a reward of 10. The learning agent is given a simple set of rules regarding its actions, it can move north, south, east, and west as long as there is not the end of the field in which case the agent is penalized and does not move. The code is made up of two classes, the environment and the agent.

Environment

The environment for the Q-Learning algorithm is a simple class that contains the playing field and rules regarding the playing field. In the constructor, a goal cell is chosen from a 20 x 20 matrix at random and a start position is chosen at random. The only method in this function is the *doStep* method, which simply accepts an action from the agent and applies it to the agent's current state returning the reward for that action. If the action performed results in a GOAL state, the agent is given a new start position to restart the learning process and improve on its understanding of the environment.

Agent

The agent in the Q-Learning algorithm is where the learning happens. There are variables in the agent code to track current q-values for states-action sets in the environment. Based on these values with a randomness variable and learning rate, the algorithm learns which paths to take and which paths to avoid. The algorithm performs a given amount of steps, each time adding knowledge about its state-action set to its base for future use. A greed value is also used in the algorithm to ensure that the agent will choose the best-known direction occasionally instead of always trying to learn more about the environment.

Results

Although it isn't entirely easy to see the outcome of the search algorithm, it is clear to see results in the data. For instance, as you increase the number of steps taken, you will notice that near the end the agent stops running into walls as it is making its way to the goal. This is again a quick demonstration of how a Q-learning algorithm works and would not be the best method to use in such a situation where there are other more efficient search methods. But when looking at this example and considering its possibilities when it comes to robotics and the ability or an agent to be placed in a new room and by negative and positive reinforcement to learn to complete tasks in the new environment, you can see the power in even the simplest of reinforced learning algorithms.

Future Work

There is going to be many advancements in the study of reinforcement learning over the next few years in fields like robotics, automated transportation, gaming, and much more. In fact, reinforcement learning is already being applied to such things as autonomous helicopter flight. An apprenticeship reinforcement learning algorithm was applied to study maneuvers done by professional helicopter pilots for a period of time. The controls were then set up to run simulations overnight to hone their skills. At the first test of the autonomous helicopters, they were performing maneuvers would not be attempted by anyone but the absolute best helicopter pilots. Even making claims that the agents were out performing the expert pilots in some cases (Abbeel, Coates and Ng). It is truly a testament to the possibilities of this growing field of study.

Conclusion

In conclusion, reinforcement learning is an interesting and cutting edge facet of artificial intelligence. There are many great agents to come from this type of learning. Learning like humans gives these agents the unique ability to solve problems that many thought would be impossible years ago. I am personally excited to see where this technology takes us in the future.

References

Abbeel, Pieter, Adam Coates and Andrew Ng. "Autonomous Helicopter Aerobatics through Apprenticeship Learning." The International Journal of Robotics Research (2010).

Hardesty, Larry. How Computers Can Learn Better. 29 5 2013. 30 7 2013 <http://www.mit.edu/newsoffice/2013/machine-learning-algorithm-outperforms-predecessors-0529.html>.

Russell, Stuart and Peter Norvig. Artificial Intelligence: A Modern Approach. Prentice Hall, 2009.

Watkins, Christopher. "Learning from Delayed Rewarrds." PhD Thesis. King's College, 1989.

Woergoetter, Florentin and Bernd Porr. Reinforcement Learnng. 2008. 30 7 2013 <http://www.scholarpedia.org/article/Reinforcement_learning#Background_and_History>.