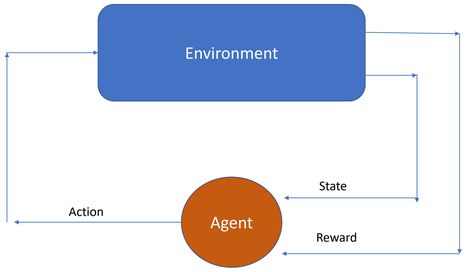
Interpretable Reinforcement Learning with Ensemble Methods

Alexander Brown and Marek Petrik
Department of Computer Science, University of New Hampshire

Motivation

- Reinforcement learning continues to set new bounds on what is possible, recently with AlphaGo and AlphaGo Zero
- Sometimes we need more than a black box
  - Healthcare – allocate scarce medical resources to handle rare ER cases
  - Targeted advertising, finance, etc.
- Prior work focuses independently on human-interpretability and reinforcement learning

Reinforcement Learning Paradigm



- A reinforcement learning problem has the form of an agent choosing actions in an environment and getting some reward
- A solution is a policy describing which action to choose in every state

Approach

- Decision trees are naturally human-interpretable
- Policy gradient descent can be used for reinforcement learning
- Classification methods can be used for policy iteration algorithms

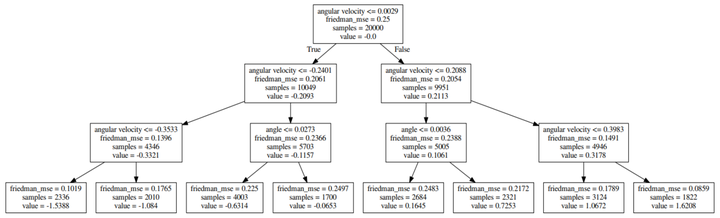


Figure 1: Regression tree from a cart-pole ensemble

Decision trees

- A decision tree represents a sequence of decisions, resulting in a prediction
- Decision trees come in two types
  - Classification trees (qualitative prediction)
  - Regression trees (quantitative prediction)
- Small decision trees are human-interpretable because a human could trace all possible decisions through the tree
- Decision trees (often regression trees) can be combined into an ensemble in various ways for a more expressive model

Benchmarks

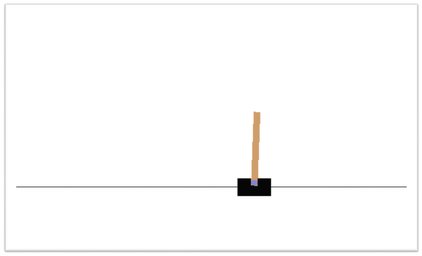


Figure 2: Cart-Pole diagram

- Cart-Pole environment
  - State: Cart position, cart velocity, pole angle, pole angular velocity
  - Actions: Move the cart left or right
  - Goal: Keep the pole upright as long as possible

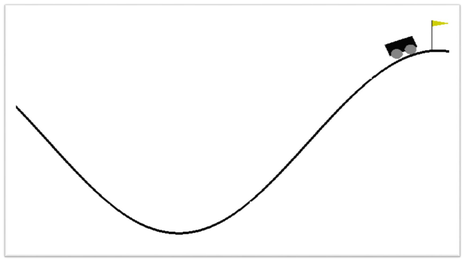
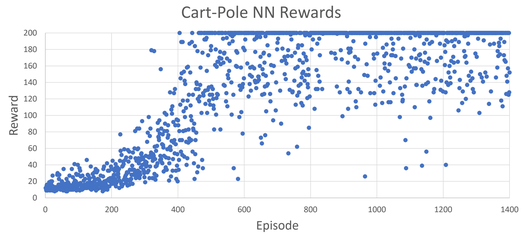


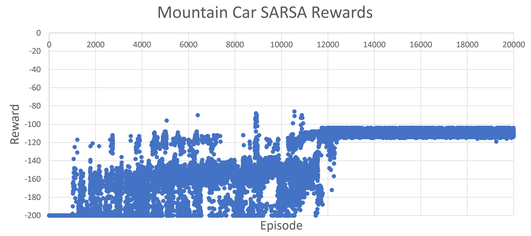
Figure 3: Mountain Car Diagram

- Mountain Car Environment
  - State: Car position, car velocity
  - Actions: Apply force forward or backward
  - Goal: Drive the car up the taller hill (the hill is too tall to drive straight up)

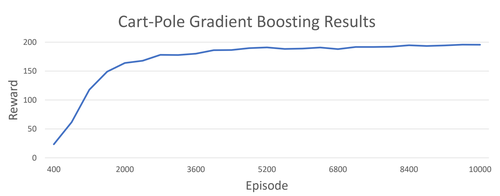
Results



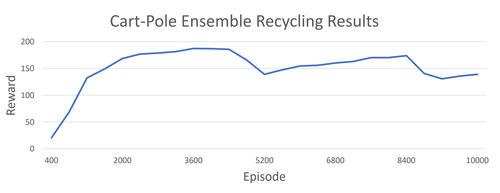
- The ensemble effectively learned the policy from the neural network (NN)



- The ensemble effectively learned the policy from SARSA



- Policy gradient boosting was able to learn in the cart-pole environment
- Unable to match performance of the neural network



- Policy gradient boosting with ensemble recycling was able to make progress in the cart-pole environment but was set back by the recycling
- Unable to match performance of policy gradient boosting without ensemble recycling

Contributions

- Implemented solution to reinforcement learning problems that humans can interpret
- Analyzed other methods of building human-interpretable solutions

Methods

- Trained ensembles of decision trees using policy data from a stronger learner
- Incrementally trained ensembles of decision trees to perform policy gradient boosting
- Periodically recycled portions of the ensemble in policy gradient boosting to reduce space and time needs, a technique we refer to as "ensemble recycling"

Further Research

- Investigate using boosted regression trees to compute solutions to more complex problems based off of another solution

Acknowledgements

We would like to thank the Hamel Center for Undergraduate Research for facilitating the SURF program, and Mr. Dana Hamel and Brad Larsen for their generosity.