Curiosity-Driven Exploration with Planning Trajectories

Tyler Streeter

PhD Student, Human Computer Interaction
Iowa State University

http://www.vrac.iastate.edu/~streeter

Overview

- Reinforcement learning (RL) agents can reduce learning time dramatically by planning with learned predictive models.
- Planning trajectories, sequences of imagined interactions with the environment, are one way to use such predictive models.
- In complex environments, planning agents need an intrinsic drive to improve their predictive models, such as a curiosity drive that rewards agents for experiencing novel states.
- Curiosity acts as a higher form of exploration than simple random action selection because it encourages targeted investigation of interesting situations.
- In a task with multiple external rewards, we show that curious planning RL agents outperform non-curious planning agents in the long run.

Agent Architecture

- The agent architecture combines several components
  - Separate feedforward neural networks for policy and value function
  - Temporal difference learning with eligibility traces
  - Planning with a feedforward neural network for sensory/reward prediction
  - Prediction uncertainty estimation with a feedforward neural network
  - Curiosity rewards based on prediction uncertainty
- An open source implementation of this architecture is available online: http://verve-agents.sourceforge.net

Main Algorithm

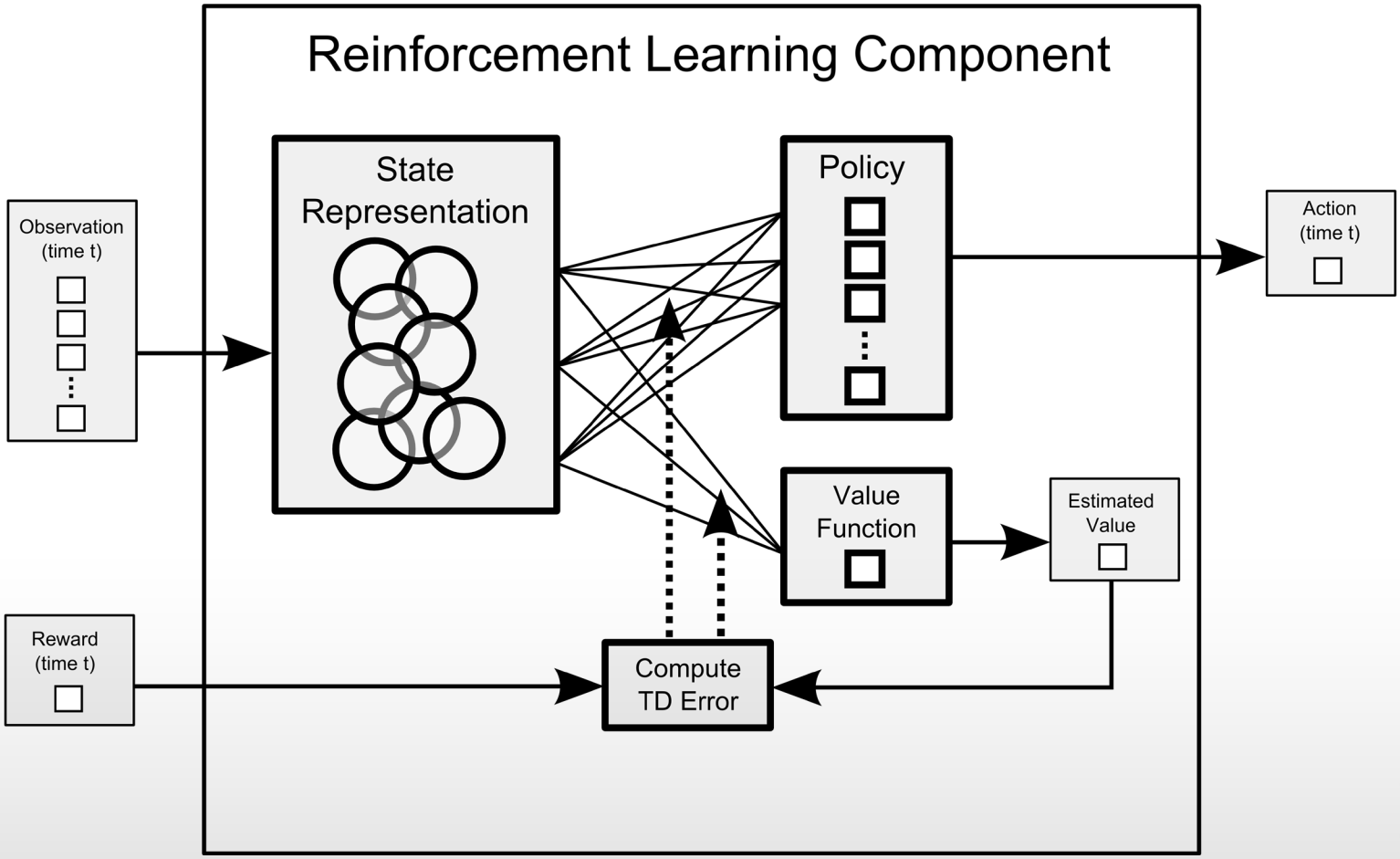
loop forever:

- based on the previous observation and action, predict the current observation and reward
- train the predictor using the actual current observation and reward
- train the uncertainty estimator using the predictor's MSE
- loop until uncertainty is too high:
  - total reward = predicted current reward + curiosity reward (proportional to uncertainty estimation)
  - perform TD learning on the value function and policy using the predicted current observation and the total reward
- use the policy to choose an action
- update the environment with the chosen action; generate the next observation and reward

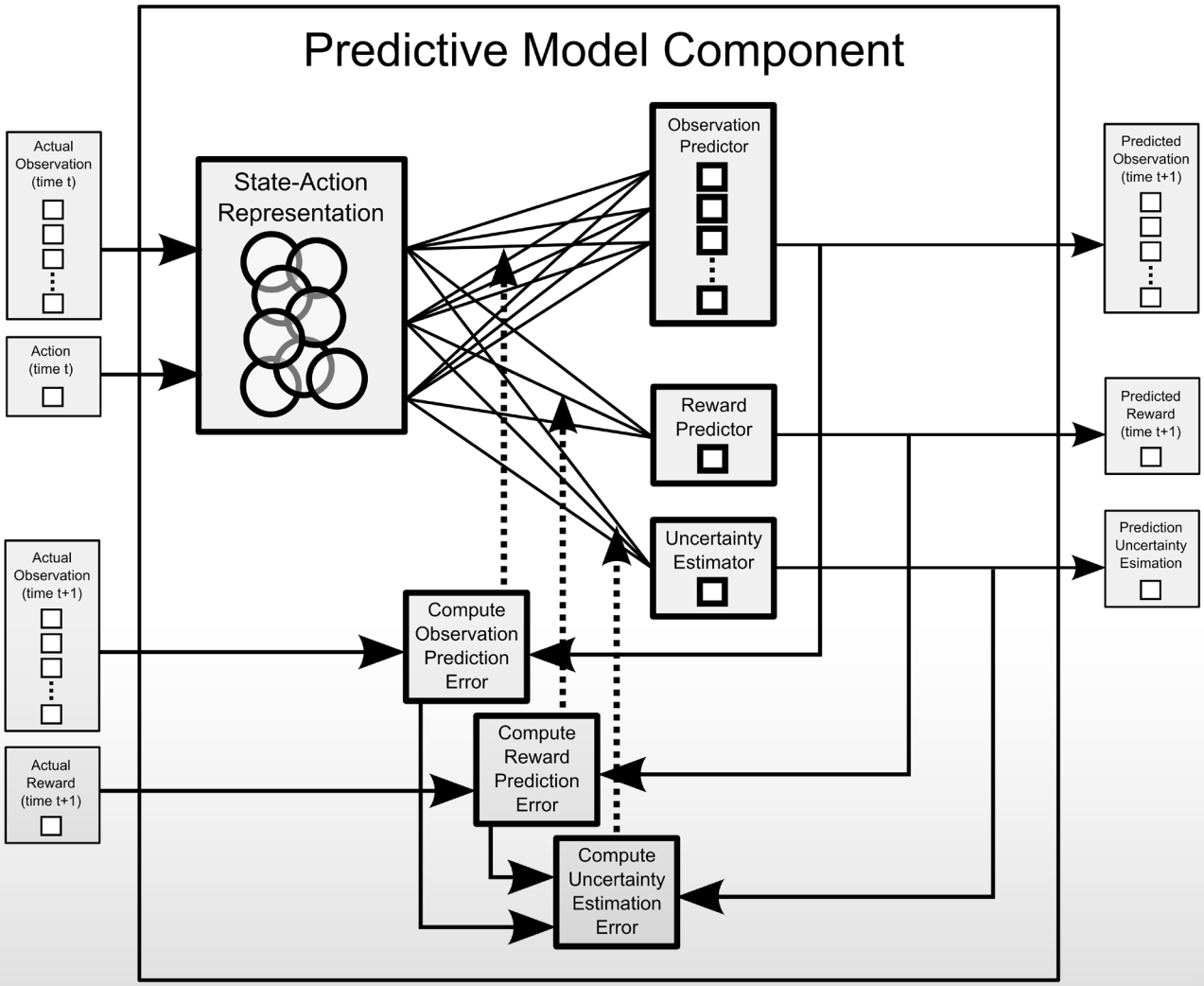
Curiosity Implementation

- Curiosity here is simply a reward proportional to prediction errors (i.e. the MSE of the model's predictions), similar to the curiosity reward used in Barto, Singh, & Chentanez, 2004.
- To generate curiosity rewards during planning (where prediction errors cannot be measured), a second predictor is required which learns to estimate the model's prediction errors.
- This curiosity implementation is sufficient for the present investigation. However, more powerful curiosity mechanisms are available which use curiosity rewards proportional to the decrease in prediction errors over time (e.g., Schmidhuber, 1991, and Oudeyer & Kaplan, 2004).

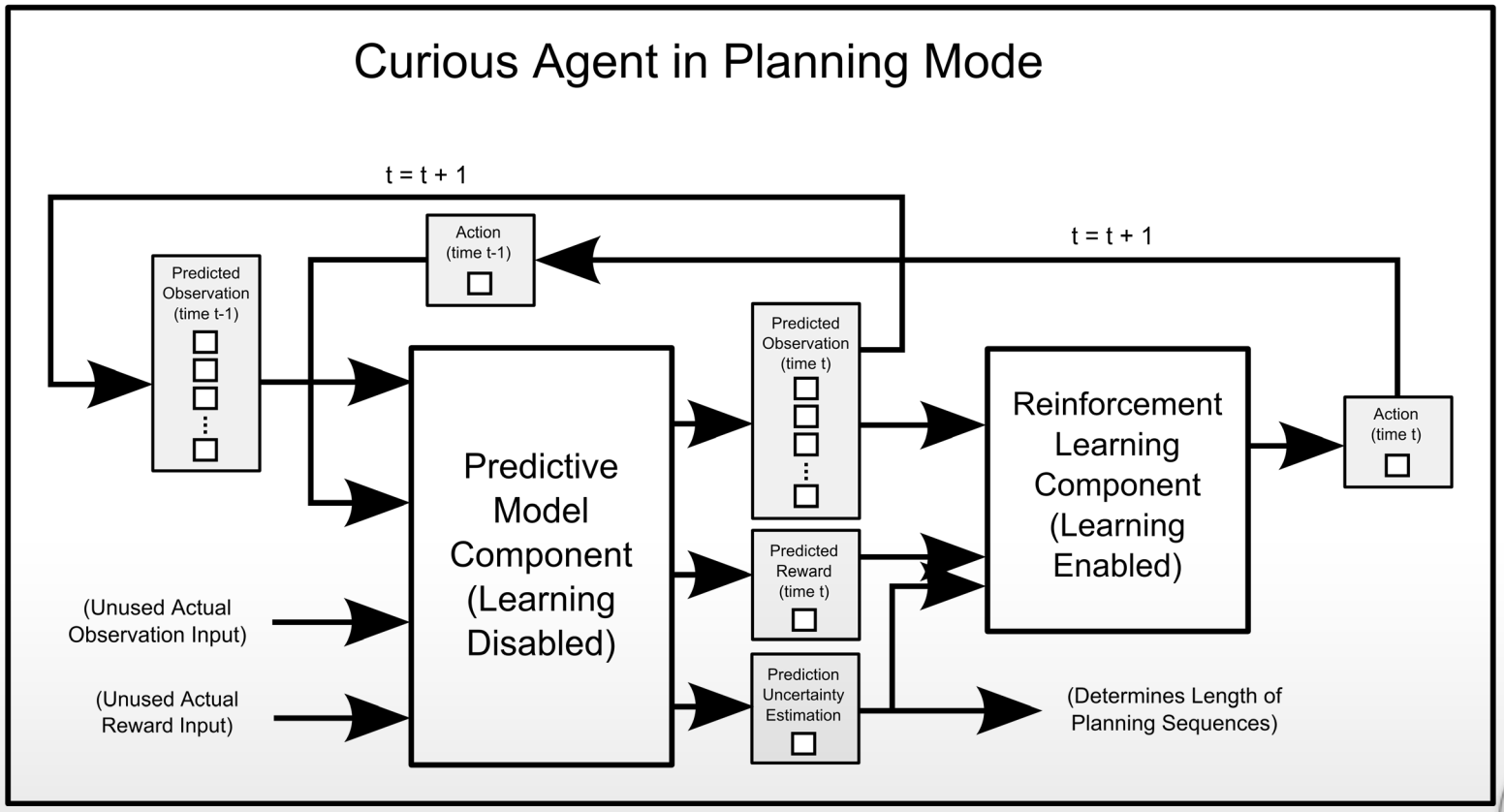
Agent Architecture



Agent Architecture

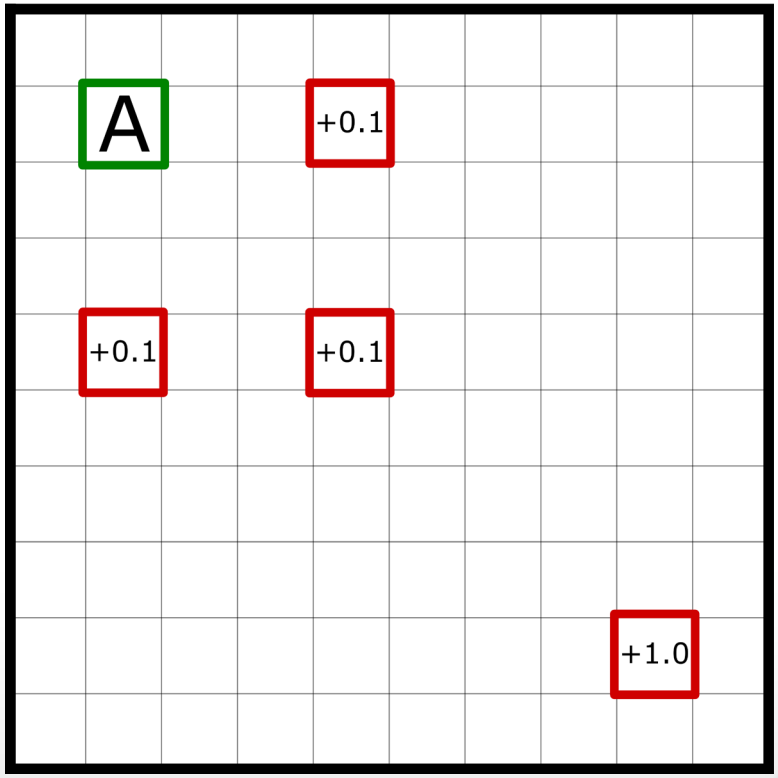


Agent Architecture

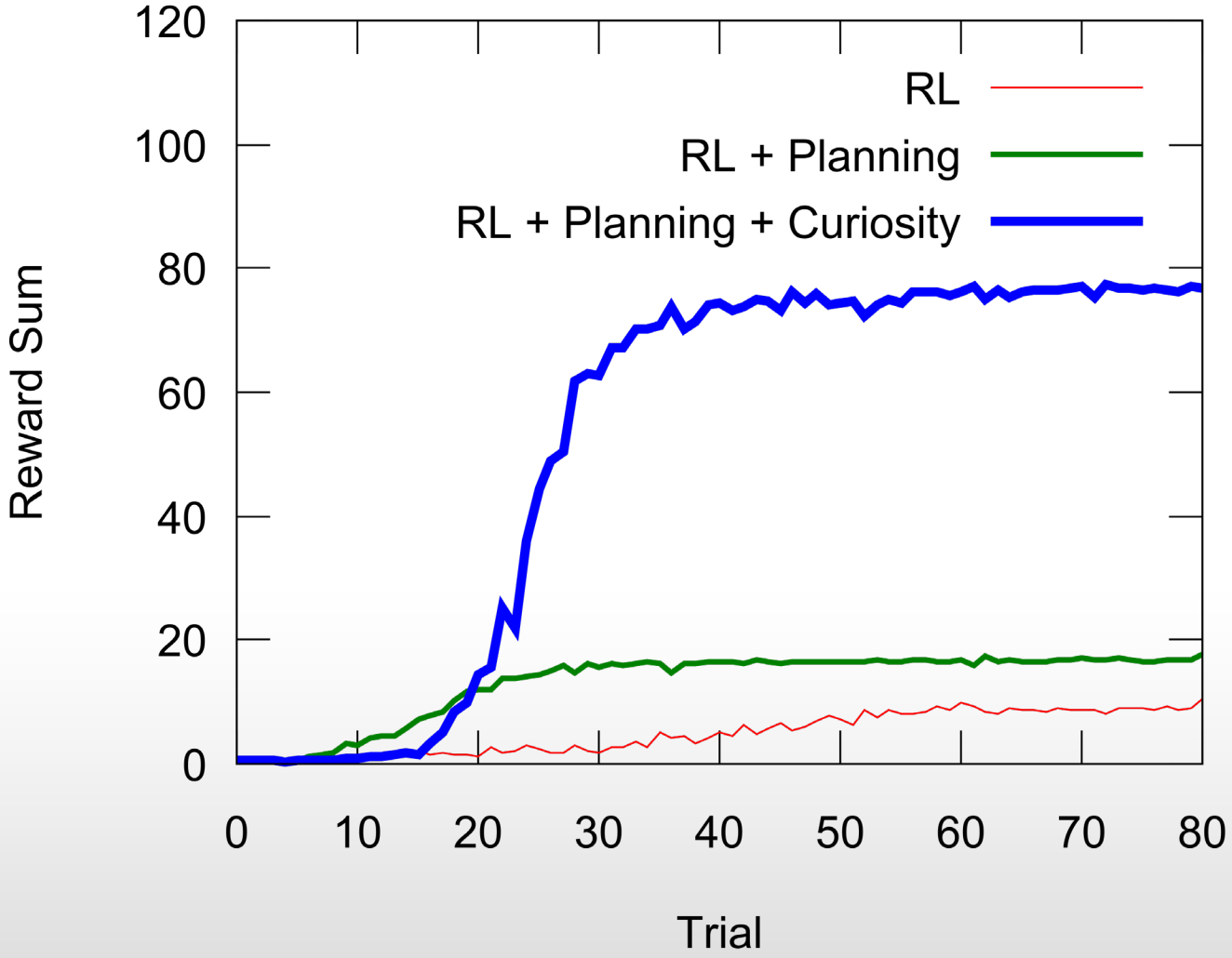


Experiment

- In this discrete environment an agent is able to sense its 2D position and can move left, right, up, down, or remain in place.
- The agent starts each trial in space ‘A’ and interacts for 100 time steps.
- There are four external rewards (three small and one large).
- Performance = sum of rewards received over the course of a single trial.



Results


120
100
80
60
40
20
0
Reward Sum
0
10
20
30
40
50
60
70
80
Trial
RL
RL + Planning
RL + Planning + Curiosity

Discussion

- The curious agent plot shows a noticeable delay before any visible improvement, indicating a curiosity-driven exploration phase. Around trial 15, the agent “runs out” of interesting situations to explore, so it begins maximizing external reward intake.
- In general, curious agents may take longer to improve performance on specific tasks, but they learn a broader knowledge base that might be applicable to a wide variety of tasks.

Future Work

- Experimentation with various curiosity mechanisms (e.g., curiosity rewards proportional to the decrease in prediction errors)
- Higher-dimensional state spaces with dimensionality reduction techniques (e.g., PCA, SOMs, ICA)

Links

- My Website
  http://www.vrac.iastate.edu/~streeter
- Verve Agent Implementation
  http://verve-agents.sourceforge.net
- Iowa State Virtual Reality Applications Center
  http://www.vrac.iastate.edu
- Iowa State HCI Graduate Program
  http://www.hci.iastate.edu

References

- Barto, A., Singh, S., & Chentanez, N. 2004. Intrinsically motivated learning of hierarchical collections of skills. In *3rd International Conference on Development and Learning.*
- Oudeyer, P.-Y., and Kaplan, F. 2004. Intelligent adaptive curiosity: a source of self-development. In Berthouze, L., et al, eds., *Proceedings of the 4th International Workshop on Epigenetic Robotics*, volume 117, 127-130. Lund University Cognitive Studies.
- Schmidhuber, J. 1991. Curious model-building control systems. In *Proceedings of the International Joint Conference on Neural Networks*, vol. 2, 1458-1463. IEEE.