

COMPSCI 295, Fall 2019

Assignment 5

Instructor: Rina Dechter

Due (by email): Friday, Nov 22

We had 4 presentations of papers in the last class. Below are a set of questions, one from each presentation. Please answer 2 of the 4 questions. You should send a single email to and to the 2 graders of the questions you answered. Precede each question with a few sentences of summary of the paper

1. (Paper: An Actor-Critic-Attention Mechanism for Deep Reinforcement Learning in Multi-view Environments, Presenter: Anna Schwarz, Email: hschwarz@uci.edu) : Describe the Attention Based RL Framework used in this paper and its training steps. In particular, what steps were taken to ensure all workers are trained?
2. (Paper: An analysis of linear models, linear value-function approximation, and feature selection for reinforcement learning, Presenter: Momoko Kono; momokok@uci.edu) In linear **model** approximation, we have transition probability matrix P_{Φ} and feature vector of a state $\Phi(s)$.
Let \vec{x} be a vector such that $\vec{x} = P_{\Phi}^T \Phi(s)$.

- (1) What is the size of \vec{x} ?
- (2) Show the i th element of \vec{x} in terms of ϕ and P_{Φ}
- (3) What does \vec{x} exactly indicate? Describe briefly.

3. (Policy Gradient Methods for Reinforcement Learning with Function Approximation: paper <https://papers.nips.cc/paper/1713-policy-gradient-methods-for-reinforcement-learning-with-function-approximation.pdf>), presenter: Bahareh Harandizadeh

Please answer the following question based on the paper and David Silver lecture on Policy Gradient (lecture 7): Consider three approaches for finding optimal policy (Value-Based, Actor Critic and Policy-Based) in my last slide, and briefly explain how each of this policy applies. Then enumerate some of the advantages and disadvantages of each policy (compare to the other ones).

4. (Paper title: An Analytic Solution to Discrete Bayesian Reinforcement Learning(<https://www.ics.uci.edu/~dechter/courses/ics-295/fall-2019/papers/icml-brl-8pages.pdf>)) Presenter: Yanqi Gu (yanqig1@uci.edu)

In section 3.1, why equation (7) could achieve exploration/exploitation tradeoff? Hint: compare with pure exploitation and check experiment results in section 5.