Performance Loss Bounds for Approximate Value Iteration with State Aggregation

Mathematics of Operations Research

Van Roy

February 18, 2019

Problem Definition

- What is this paper about?
  How to approximate J^*, the optimal cost-to-go function (value function), with Φr?
  - State Aggregation
  - Value Iteration
  - In particular, how to find $r \in \Re^K$?
- What is this paper NOT about?
  How to partition the space (or find $\Phi \in \Re^{|\varphi| \times K}$)?

[0] $|\varphi|$ is the number of states

Approximating J

- $$\min_r \|J - \Phi r\|_{2,\pi} \tag{1}$$
- Projection with respect to a weighted Euclidean norm $\|.\|_\pi$

$$\|J\|_{2,\pi} = \left(\sum_{x \in \varphi} \pi(x) J^2(x) \right)^{1/2}$$

$\pi \in \Re_+^{|\varphi|}$ is a vector of weights, showing the *importance* of each state

- To get ***r***, we need to run value iteration

$$\Phi r^{(l+1)} = \Pi_\pi T \Phi r^{(l)}$$

Calculating r

$$\Phi r^{(l+1)} = \Pi_\pi T \Phi r^{(l)}$$

- T is the dynamic programming operator
- T is a contraction
- Π_π is max-norm nonexpansive
- Π_π operator (matrix) projects onto the column space of Φ with respect to a weighted Euclidean Norm that minimizes equation (1)
- $$\Pi_\pi = \Phi(\Phi^T D \Phi)^{-1} \Phi^T D$$

  where $D = diag(\pi_i)$

Bounds

- Without considering the importance weights, and when μ is greedy with respect to $\tilde{J}$ we have the bound[1]:

$$\|J_\mu - J^*\|_\infty \leq \frac{2\alpha}{1-\alpha}\|J^* - \tilde{J}\|_\infty \qquad (2)$$

- Considering the importance weights and when $\Phi\tilde{r} = \Pi_\pi T\Phi\tilde{r}$
  - *Approximation Error Bound*

$$\|\Phi\tilde{r} - J^*\|_\infty \leq \frac{2}{1-\alpha}\min_{r\in\Re^K}\|J^* - \tilde{J}\|_\infty$$

  - *Performance Loss Bound*

$$(1-\alpha)\|J_{\mu\tilde{r}} - J^*\|_\infty \leq \frac{4\alpha}{1-\alpha}\min_{r\in\Re^K}\|J^* - \Phi r\|_\infty$$

  - HOWEVER!!!

[1] $\|J_\mu - J^*\|_\infty$ is performance loss

Using the Invariant Distribution

- if $\pi_{\tilde{r}}$ is is the invariant state distribution of transition matrix $P_{\mu_{\tilde{r}}}$ [2]

$$(1-\alpha)\pi_{\tilde{r}}^T(J_{\mu\tilde{r}} - J^*) \leq 2\alpha \min_{r\in\Re^K} \|J^* - \Phi r\|_\infty$$

- Compare it with:

$$(1-\alpha)\|J_{\mu\tilde{r}} - J^*\|_\infty \leq \frac{4\alpha}{1-\alpha} \min_{r\in\Re^K} \|J^* - \Phi r\|_\infty$$

- Weighting Euclidean Norm projection by the invariant distribution of a greedy (or ϵ-greedy) policy improves the performance loss.

[2] π is an invarient of P if $\pi^T P = \pi^T$

Thank You!