Analyzing Feature Generation for Value-Function Approximation

Ronald Parr, Christopher Painter-Wakefield  
Duke University  
Lihong Li, Michael Littman  
Rutgers University

**Value-Function Approximation in RL**

**Standard Bellman equation:**

\[ V^\pi = R + \gamma PV^\pi \]

**T = Bellman operator**

**Linear value-function approximation:**

\[ \hat{V} = \sum_i w_i \phi_i \quad \text{linear combination of features} \]

\[ \hat{V} = \Phi W = \Pi T \hat{V} \]

**Fixed point**

\[ \Pi = \text{Projection into span of } \Phi = [\phi_1, \phi_2] \]

---

**Feature Generation in RL**

**Features = Basis functions for linear value function approximation**

- **Mahadevan & Maggioni [05]**  
  - Use connectivity graph of the state space  
  - Use spectral clustering/ manifold learning techniques

- **Keller, Mannor & Precup [06]**  
  - Use the Bellman error  
  - Function approximation applied to the Bellman error

**Bellman Error:**

\[ BE(\hat{V}) = TV - V \]

---

**Theoretical Result 1: Approximation Bounds**

Linear fixed point approximation error [Van Roy 98]:

\[ \left\| V^* - \hat{V} \right\| \leq \frac{1}{\sqrt{1 - \gamma}} \left\| V^* - \Pi V^* \right\| \]

\[ \Pi V^* \]

**Our first theorem** says that adding one BEBF to our basis improves our approximation bound at least as much as one step of value iteration:

Let \( \hat{V} \) be the linear fixed point solution using a sequence of normalized BEBFs \( \phi_1, \ldots, \phi_n \). If \( \left\| V^* - \hat{V} \right\| \leq x \), then for new BEBF \( \phi_{n+1} \) with \( \Phi = [\Phi, \phi_{n+1}] \), and corresponding \( \Pi \), the improvement in the approximation bounds is

\[ \left\| V^* - \Pi V^* \right\| < \left\| V^* - T \hat{V} \right\| \leq x \]

**Our second theorem** says that we can use an approximation, \( \hat{V} \), to the Bellman error as long as the approximation is reasonably accurate:

If (1) the angle between \( \phi_i \) and \( \hat{V} \) is less than \( \cos^{-1}(\gamma) \) radians and (2) \( \hat{V} \neq V^* \), then there exists a \( \beta \) such that \( \left\| V^* - (\hat{V} + \beta \phi_i) \right\| < \left\| V^* - \hat{V} \right\| \). Moreover, if conditions (1) and (2) hold and \( \phi_i \) is not in the span of \( \Phi \), then for \( \Phi' = [\Phi, \phi_{n+1}] \) and corresponding \( \Pi' \),

\[ \left\| V^* - \Pi' V^* \right\| < \left\| V^* - \Pi V^* \right\| \]

**Theoretical Challenge**

**Can we generate new basis functions with provable quality guarantees?**

**Our Answer:** Yes — If we can approximate the Bellman error with sufficient accuracy

\[ \phi_{n+1} = BE(\hat{V}) = TV - \hat{V} \]

\[ \Phi' = [\phi_1, \phi_2, \ldots, \phi_n, \phi_{n+1}] \]

**Application to the 50-state Chain Problem**

[lagoudakis & Parr 03]

- 50 States numbered 1-50  
- Noisy Actions: Move Right, Move Left (0.9 success)  
- +1 Rewards at 10, 41  
- 0.8 Discount

**Empirical Results**

**Application to PuddleWorld**

From Boyan & Moore [1994]

“Ground Truth”

**Results Using Exact BEBFs w/model**

Error vs. Number of BEBFs w/8,000 samples, approximate BEBFs

**LISPI w/BEF, LWR Bellman Error Approximation**

All Features Generated Automatically

**1 Training episode = 10 steps or until goal is reached**

---

**From Boyan & Moore [1994]**

LWR = Locally Weighted Regression

LSTD w/LWR

LSTD w/LWR