Methods for Bellman Error Basis Function construction

Doina Precup, School of Computer Science, McGill University
Collaboration with A-m. Farahmand, M. M. Fard, Y. Grinberg & J. Pineau

Function approximation is crucial for obtaining good results in large reinforcement learning tasks, but the problem of devising a good function approximator is difficult and often solved in practice by hand-crafting the “right” set of features. In the last decade, a considerable amount of effort has been devoted to methods that can construct value function approximators automatically from data. Among these methods, Bellman error basis function construction (BEBF) are appealing due to their theoretical guarantees [6] and good empirical performance in difficult tasks [4]. In this work, we discuss on-going developments of methods for BEBF construction based on random projections [2] and orthogonal matching pursuit [1].

Bellman Error Basis Function construction

Let $\langle S, A, R^a, P^a, \gamma \rangle$ be a Markov Decision Process (MDP) where R^a andP^a represent the reward vector and probability matrix for each action a. Let $TV = \max_a | R^a + \gamma P^a V$ be the Bellman operator.

Let Φ be a feature matrix of size $n \times k$, where n is the number of states in S and k is the number of features. A linear approximation to the value function V can be defined as: $\hat{V} = \Phi \mathbf{w}$, where $\mathbf{w}$ is a parameter vector to be estimated from data. Linear fixed-point methods obtain $\mathbf{w}$ by solving a fixed-point equation: $\hat{V} = \Phi \mathbf{w} = \Pi \hat{V}$ where Π is a projection operator on the span of Φ, typically weighted by a stationary distribution of state visitations (which gives implicitly an importance to each state).

BEBF construction starts with a small number of features (eg $k = 1$) and gradually adds feature vectors to Φ so as to minimize the Bellman error. Exact BEBF construction would add the feature $\phi_{k+1} = T\hat{V} - \hat{V}$, which perfectly approximates the error. However, in practice one needs to pick an approximation to ϕ_{k+1} which is easy to compute.

Random projections for BEBF

We developed an algorithm that uses the idea of applying random projections specifically in very large and sparse feature spaces (e.g. 10^6 dimensions) [2]. The idea is to iteratively project the original features into exponentially smaller-dimensional spaces. Then, we apply linear regression in these spaces, using temporal difference errors as targets, in order to approximate BEBFs. A finite-sample analysis provides information on how to determine the sizes to be used for the projection.

This method compares favourably, from a computational point of view, to other feature extraction methods in high dimensional spaces, as each iteration takes only polylogarithmic time in the number of dimensions. Because we use agnostic random projections, minimal domain knowledge is needed.

Value Pursuit Iteration

An alternative way to approach the problem of basis function construction is to have a large dictionary of features, but to select just a few of them for the value function representation, by using a strong model selection procedure, such as orthogonal matching pursuit (OMP) [3, 5]. Value pursuit iteration [1] is an approximate value iteration algorithm which uses OMP at its core in order to find a good sparse approximation of the optimal value function given a set of features. However, the algorithm also extends the set of basis functions after each iteration, based on the currently learned value function. A finite-sample analysis of the algorithm shows that increasing the size of the dictionary can lead to much smaller approximation errors.

References

[1] Amir Massoud Farahmand and Doina Precup. Value pursuit iteration. In *NIPS*, pages 1349–1357, 2012.

[2] Mahdi Milani Fard, Yuri Grinberg, Amir Massoud Farahmand, Joelle Pineau, and Doina Precup. Bellman error based feature generation using random projections on sparse spaces. ArXiv CoRR abs/1207.5554, 2012.

[3] Jeff Johns. *Basis Construction and Utilization for Markov Decision Processes using Graphs*. PhD thesis, University of Massachusetts, Amherst, 2010.

[4] Philipp W. Keller, Shie Mannor, and Doina Precup. Automatic basis function construction for approximate dynamic programming and reinforcement learning. In *Proceedings of ICML*, 2006.

[5] Christopher Painter-Wakefield and Ronald Parr. Greedy algorithms for sparse reinforcement learning. In *Proceedings of ICML*, 2012.

[6] Ronald Parr, Christopher Painter-Wakefield, Lihong Li, and Michael L. Littman. Analyzing feature generation for value function approximation. In *Proceedings of ICML*, 2007.