Efficient Learning and Planning with Compressed Predictive States

William Hamilton, Mahdi Milani Fard, Joelle Pineau

McGill
School of Computer Science
Montreal, Canada

Problem

Goal is to construct RL agents capable of *agnostic* learning, i.e. learning and planning in complex partially observable domains without any prior knowledge (no known state-space, transition probabilities etc.).

Contribution

Developed a novel, efficient model-based RL algorithm for *agnostic* learning and planning. The learning algorithm combines predictive state representations (PSRs) [1] and random projections [5] in order to:

- Regularize the model.
- Increase computational efficiency of learning.

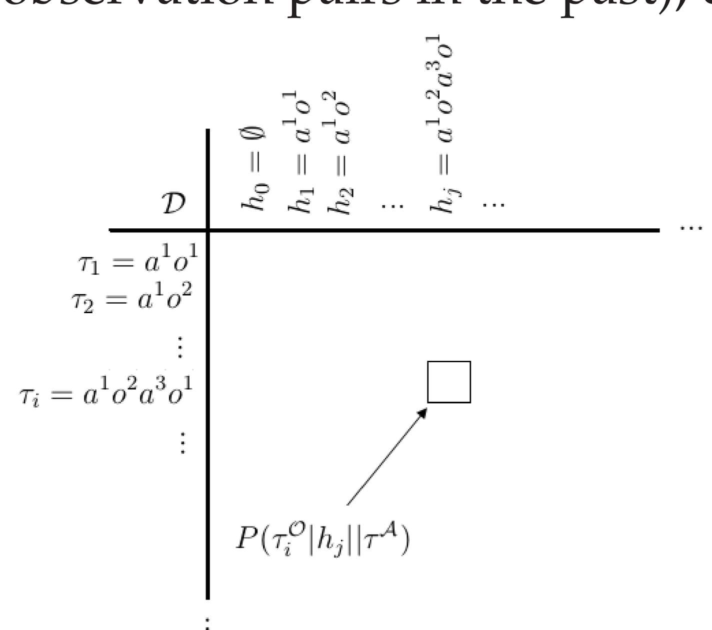
The planning algorithm uses a modified version of fitted-Q value iteration [3].

PSR Models

Predictive state representations model dynamical systems using only observable events [1].

- More general than hidden state learning approaches (e.g. EM with HMMs).
- Immune to local minima (e.g. TPSRs [2]).

Goal is learning the conditional probabilities of tests (sequences of action-observation pairs in the future), denoted τ_i, given histories (sequences of action-observation pairs in the past), denoted h_j.



Exploiting Sparsity

We say $\mathcal{D}$ is k sparse if $\forall \mathbf{c_i} \in \mathcal{D} : ||\mathbf{c_i}||_0 \leq k$.
Only a subset of tests possible given any history.
With sparsity condition, most of the information is preserved after **random projections**.
We compress $\mathbf{Y}^{m \times n}$ to $\mathbf{X}^{d \times n}$, where $d \ll m$:

$$\mathbf{X} = \Phi \mathbf{Y}$$

$\Phi^{d \times m}$ is a Johnson-Lindestrauss projection.

CPSR Learning Algorithm

Let $\mathcal{T}$ denote the set of all defined tests and $\mathcal{H}$ the set of all possible histories. Define the *observable matrices*:

- $\mathcal{P}_{\mathcal{T},\mathcal{H}}$: joint probability of all tests $\tau_i \in \mathcal{T}$ and histories $h_j \in \mathcal{H}$.
- $\mathcal{P}_{\mathcal{T},ao,\mathcal{H}}$: joint probability of tests $\tau_i \in \mathcal{T}$ and histories $h_j \in \mathcal{H}$ with action-observation pair ao prepended to each test.
- $\mathcal{P}_{\mathcal{H}}$: marginal probability for all $h_j \in \mathcal{H}$.

Obtain **compressed estimates** of the observable matrices: $\Phi_{\mathcal{T}} \hat{\mathcal{P}}_{\mathcal{T},\mathcal{H}} \Phi_{\mathcal{H}}^T$, $\hat{\mathcal{P}_{\mathcal{H}}} \Phi_{\mathcal{H}}^T$, $\Phi_{\mathcal{T}} \hat{\mathcal{P}}_{\mathcal{T},ao,\mathcal{H}} \Phi_{\mathcal{H}}^T$.

↓

Use **regression** on the **compressed estimates** to build **compact model**:

$$\mathbf{c_1} = (\Phi_{\mathcal{T}} \hat{\mathcal{P}}_{\mathcal{T},ao,\mathcal{H}} \Phi_{\mathcal{H}}^T) \mathbf{1}_d$$

$$\mathbf{C}_{ao} = (\Phi_{\mathcal{T}} \hat{\mathcal{P}}_{\mathcal{T},ao,\mathcal{H}} \Phi_{\mathcal{H}}^T)(\Phi_{\mathcal{T}} \hat{\mathcal{P}}_{\mathcal{T},\mathcal{H}} \Phi_{\mathcal{H}}^T)^+ \ \forall ao$$

$$\mathbf{c}_\infty = \hat{\mathcal{P}_{\mathcal{H}}} (\Phi_{\mathcal{T}} \hat{\mathcal{P}}_{\mathcal{T},\mathcal{H}} \Phi_{\mathcal{H}}^T)^+$$

- $\mathbf{c_1}$ is the initial **predictive model state**.
- $\mathbf{C}_{ao}$ are **update operators**.
- $\mathbf{c}_\infty$ is a **normalizer**.

Fitted-Q Planning

Define the **action-value (Q) function**:

$$Q : \mathbf{C} \times \mathcal{A} \rightarrow \mathbb{R}$$

$\mathbf{C}$ is the space of CPSR model states and $\mathcal{A}$ the set of actions. $Q(\mathbf{c}, a)$ is expected return given by taking action $a \in \mathcal{A}$ in predictive state $\mathbf{c} \in \mathbf{C}$. We estimate $\hat{Q}(\mathbf{c}, a)$ iteratively [3].

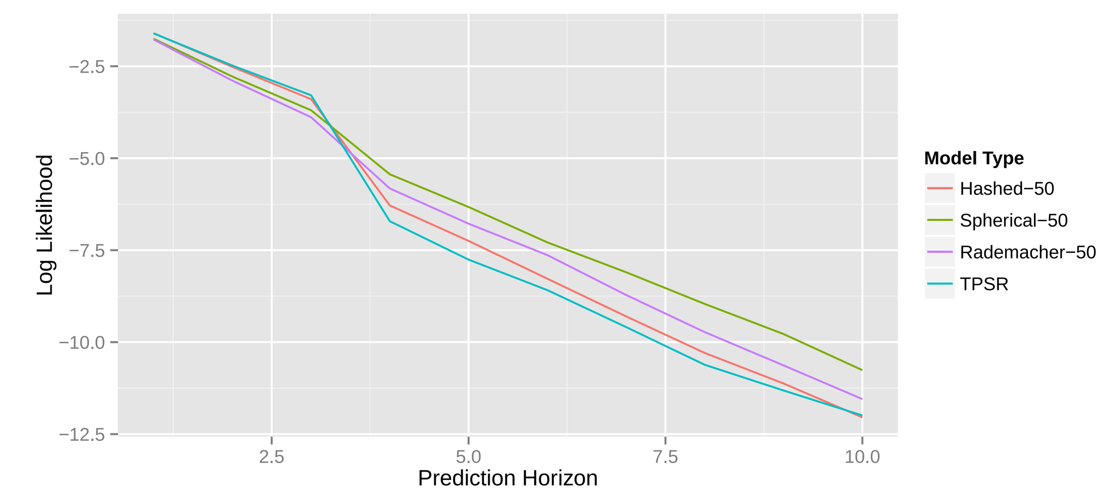
Build train set, $\mathbb{T} = \{(y^l, i^l), l = 1, ..., |\mathcal{Z}|\}$. $i^l = (\mathbf{c}_t^l, a_t^l)$, $y^l = r_t^l + \gamma \max_a \hat{Q}_{k-1}(\mathbf{c}_t^l, a)$.

↓ ↑

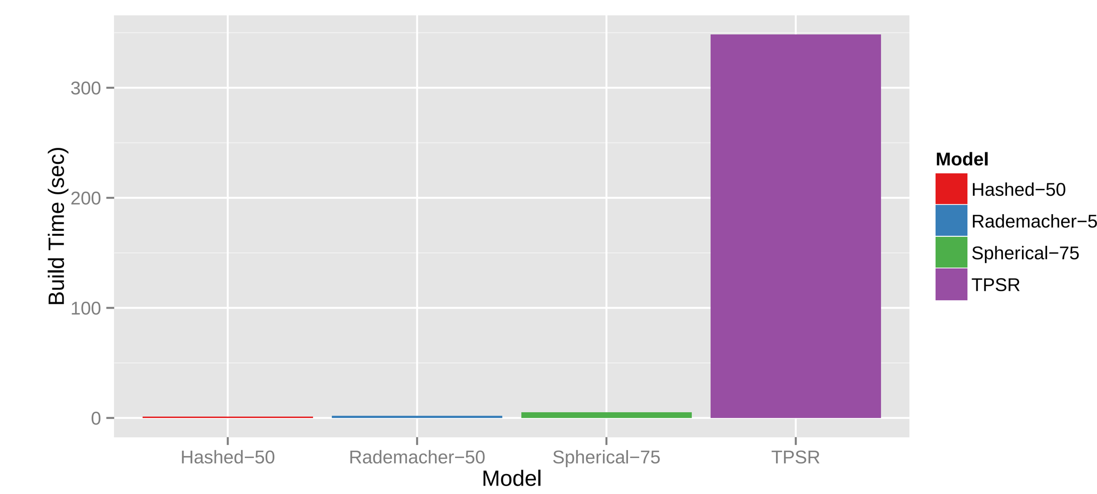
Use random forests on $\mathbb{T}$ to obtain $\hat{Q}_k$.

Empirical Results

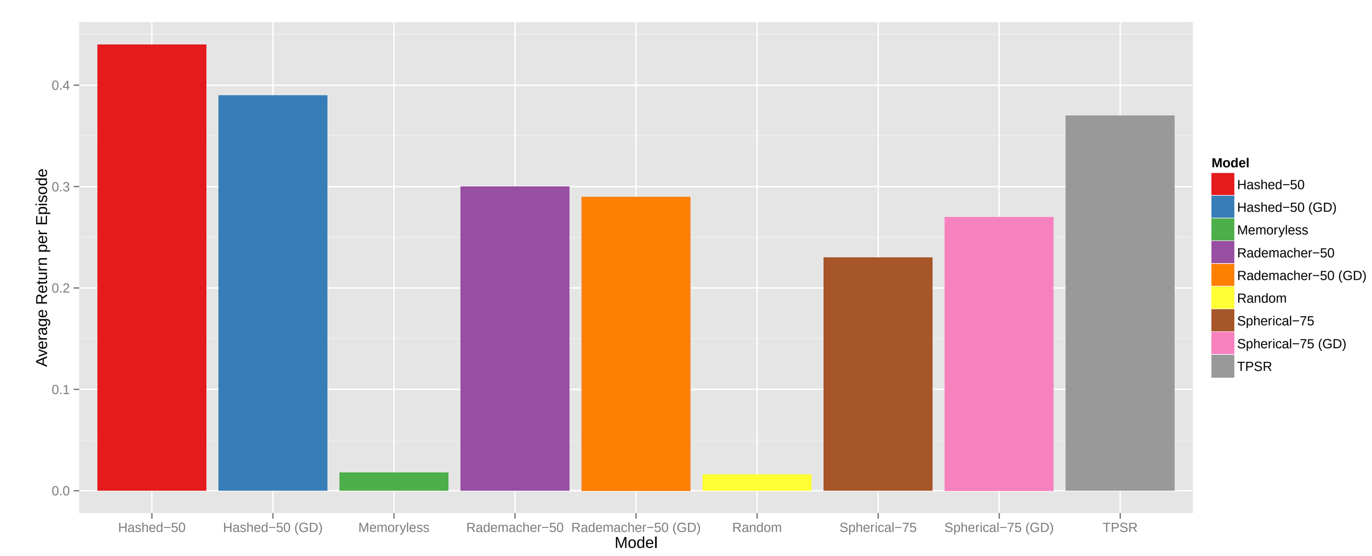
We examine the performance of different CPSRs and some baselines on two domains: *ColoredGridWorld*, a 47 state maze with 4 (noisy) actions and 81 observations; and *S-PocMan*, a partially observable variant of the video-game PacMan with approximately 10^{56} states, 4 actions, and 2^8 observations [4].



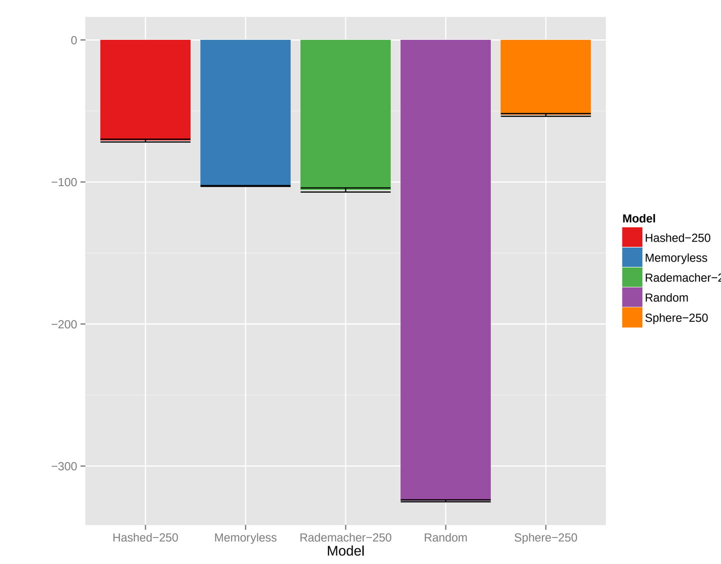
Model quality of compressed and uncompressed PSRs on *ColoredGridWorld* domain.



Empirical runtimes for constructing model of *ColoredGridWorld* domain



Average return achieved by planners in the *ColoredGridWorld* domain using uncompressed and compressed PSRs, as well as results from a memoryless controller baseline.



Average return achieved by planners in the *S-PocMan* domain.

Theoretical Results

With probability no less than $1 - \delta$ we have:
$\|\mathbf{C}_{ao}(\Phi \mathcal{P}_{\mathcal{Q},h}) - \Phi \mathcal{P}_{\mathcal{Q},ao,h}\|_{\rho(\mathbf{x})}$
$\leq \sqrt{d} \epsilon(|\mathcal{H}|, |\mathcal{Q}|, d, L_{ao}, \sigma_{ao}^2, \delta/d)$
where L_{ao} is a function of the solution's norm and σ_{ao} the induced noise after projection

$\Leftrightarrow$

With **high probability** the compressed model has **error proportional to** $\sqrt{d}$ (where d is the compressed dimension) times the usual error of performing **compressed regression** (which is bounded [5]).

Discussion

- Accuracy of CPSRs competitive with uncompressed PSRs.
- Compressed learning has far lower computational cost and regularizes the learned model.
- CPSRs facilitate model-based RL in complex partially observable domains that are intractable for uncompressed PSR learning.

References

1. M. Littman, R. S. Sutton, and S. Singh. Predictive representations of state. *NIPS 2002*.
2. B. Boots, and G. Gordon. An online spectral learning algorithm for partially observable dynamical systems. *AAAI 2011*.
3. D. Ernst,P. Geurts, L. Wehenkel, and L. Littman. Tree-based batch mode reinforcement learning. *JMLR 2005*.
4. D. Silver, and J. Veness. Monte-Carlo Planning in Large POMDPs. *NIPS 2010*.
5. O.A. Maillard, and R. Munos. Compressed least-squares regression. *NIPS 2009*.

Funding: NSERC Discovery and CGS-M grants.

Much thanks to: Yuri Grinberg, Sylvie Ong, and Doina Precup