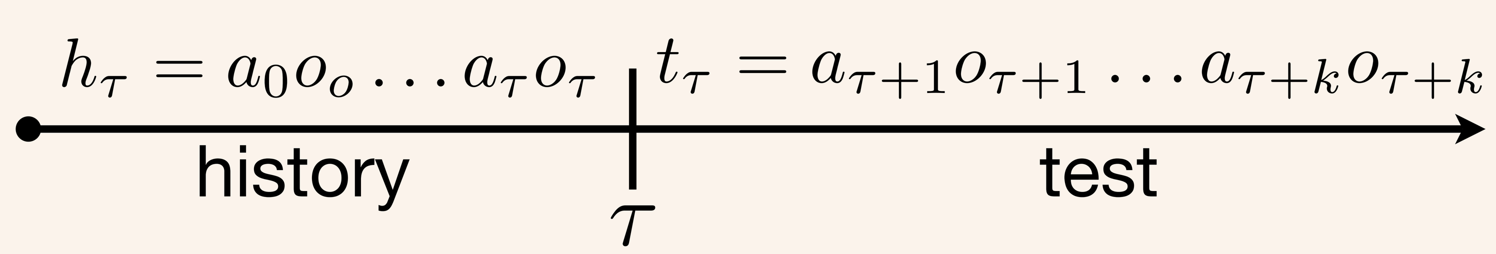
LEARNING APPROXIMATE REPRESENTATIONS OF PARTIALLY OBSERVABLE SYSTEMS

Monica Dinculescu, McGill University



Prediction of a test: $p(t|h)$

Goal: build a “useful” approximation of the system using the agent’s **interest** in the world

Test probe: $f : \mathcal{O}^* \rightarrow \mathbb{R}$

- which tests the agent wants to predict
- prediction of a test given f

$$p_f(t|h) = p(t|h) f(obs(t))$$

History probe: $g : (A \times O)^* \rightarrow \mathbb{R}$

- information available to the agent from its past
- should have enough information to predict well tests of interest
- eg. $g(h_\tau) = \sum_{i \leq \tau} \gamma^{\tau - i} (\mathbf{1}_{\mathbf{o_i}} a + (1 - \mathbf{1}_{\mathbf{o_i}}) b)$

(**1** indicates if o_i is a history feature, *a,b* are parameters, γ is discount factor)

Approximate Agent State Representation

Input: A set of data D, probes $f, g, \epsilon_f, \epsilon_g$

1. Use data to estimate $p(t|h)$ (eg. through counting)
2. Compute $p_f(t|h)$
3. **Cluster tests** using f

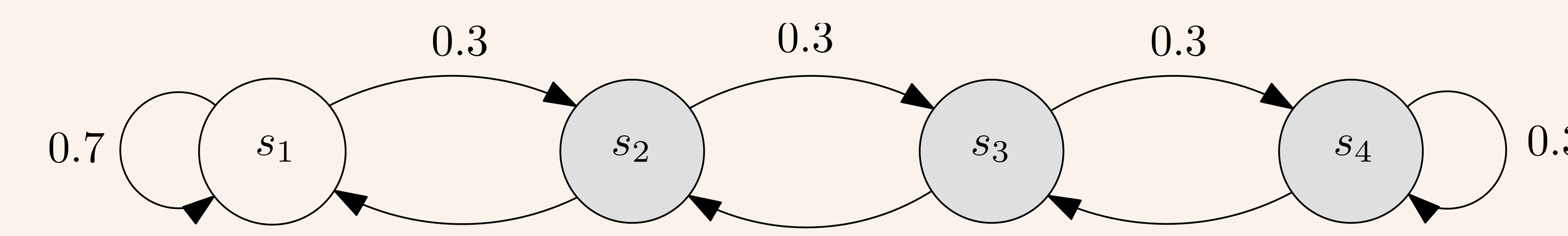
$$T' = \{[t] \mid t_1, t_2 \in [t] \Rightarrow \forall h \in H$$
$$|p_f(t_1|h) - p_f(t_2|h)| \leq \epsilon_f$$

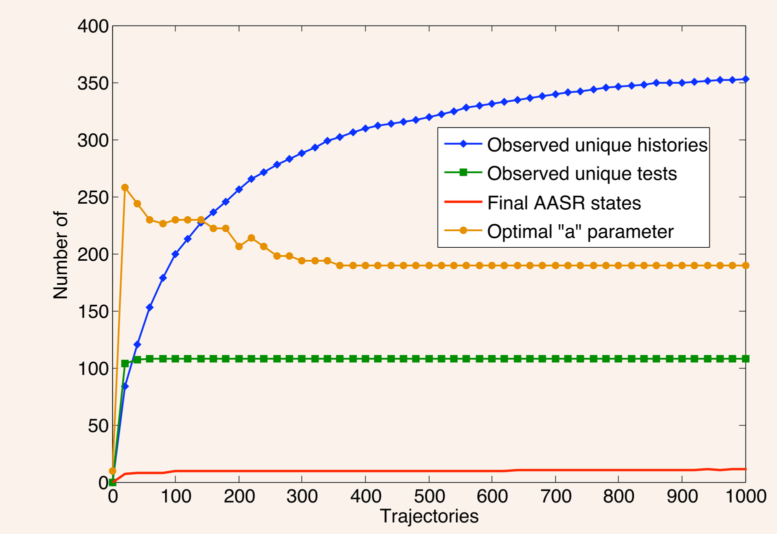
4. **Cluster histories** using g

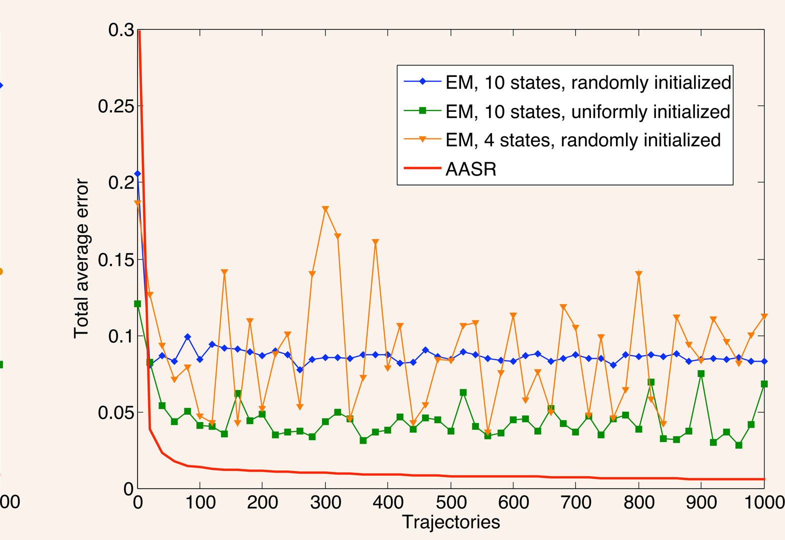
$$H' = \{[h] \mid h_1, h_2 \in [h] \Rightarrow$$
$$|g(h_1) - g(h_2)| \leq \epsilon_g$$

Output: H’, T’ and the associated prediction values $p_f(t'|h')$

Example - Tunnel World

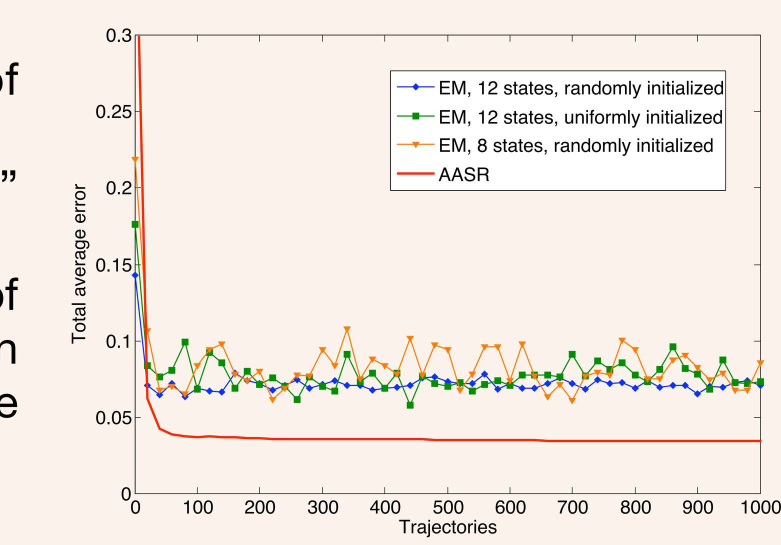






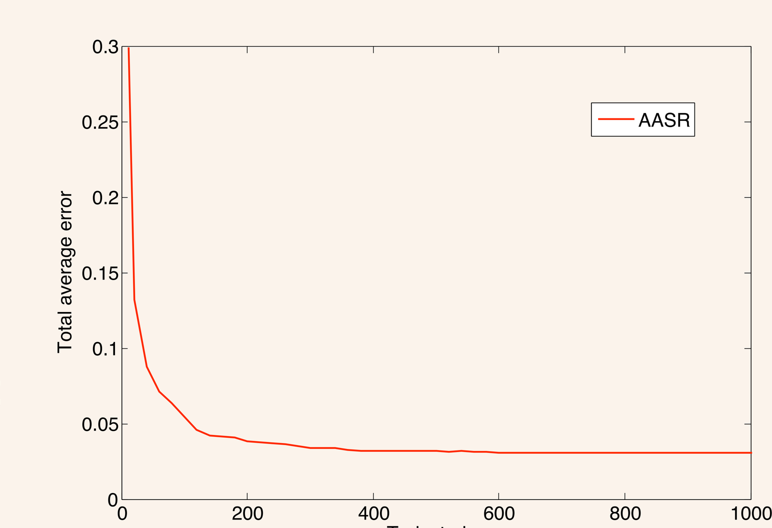
Non-Markovian

- latent variable: “direction” of motion
- direction flips when you “bounce” off the edge of the tunnel
- transitions: 0.7 in the direction of motion, 0.3 in the opposite (when the direction is left, transitions are like above)



Continuous

- start randomly in [0,1]
- with p = 0.7, transition from x to $x' = x - \delta + \epsilon_{noise}$
- otherwise, transition to $x' = x + \delta + \epsilon_{noise}$ where $\delta = 0.25$
- states that are $\leq$ 0.25 see light; otherwise, they see dark



We provide a novel approach for learning an approximate model of a partially observable environment from data. The model abstracts away the unnecessary details of the observed history and focuses only on making certain predictions of interest. The ideas we present apply in non-Markovian environments, as well as systems with continuous internal states. We illustrate our result on a small computational example.