SOME THOUGHTS ON CAUSALITY AND MACHINE LEARNING

Tom Dietterich (Oregon State)

some work joint with
George Trimponias, Zhitang Chen (Huawei Noah’s Ark Lab)

Plan

- Present an example of how causal modeling can help reinforcement learning

- Speculate the role of causal modeling in machine learning

Causal Modeling to Remove Exogenous Variables in Reinforcement Learning

- Consider training your car to drive you to work every day
- MDP
  - states: car location + traffic
  - actions: turns to make
  - cost: total time to reach the office
- Problem:
  - Your actions only control part of the cost. Most of the cost is determined by what other drivers are doing

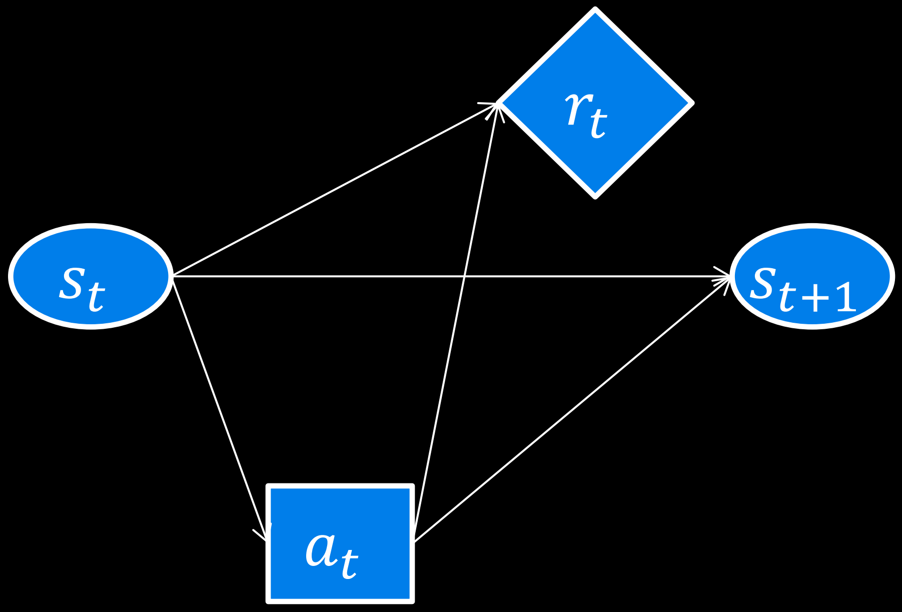
Consequences

- The cost of any policy π will have high variance
- This makes it hard to compare two policies (or to search for good policies)
  - Smaller learning rates
  - Larger sample sizes

Causal Argument

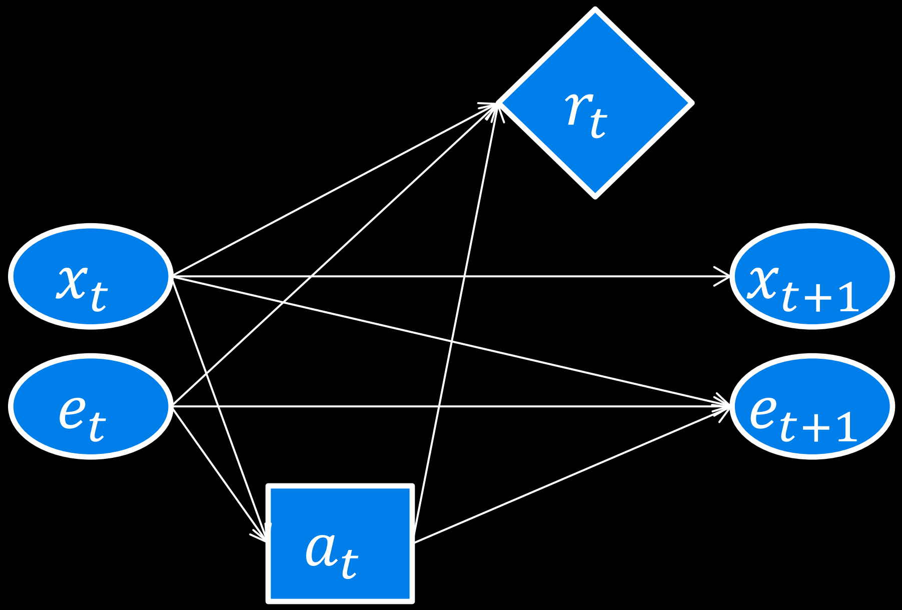
- We want to isolate the component of the reward that is caused by our actions: R_{end} ("endogenous")
- Then create an RL algorithm to find a policy π that optimizes $\mathbb{E}[\sum_t \gamma^t r_{end}(t)]$

Standard MDP Causal Diagram



Exogenous State MDP Causal Diagram

- MDP state is partitioned into $s = (x, e)$, where x is exogenous and e is endogeneous



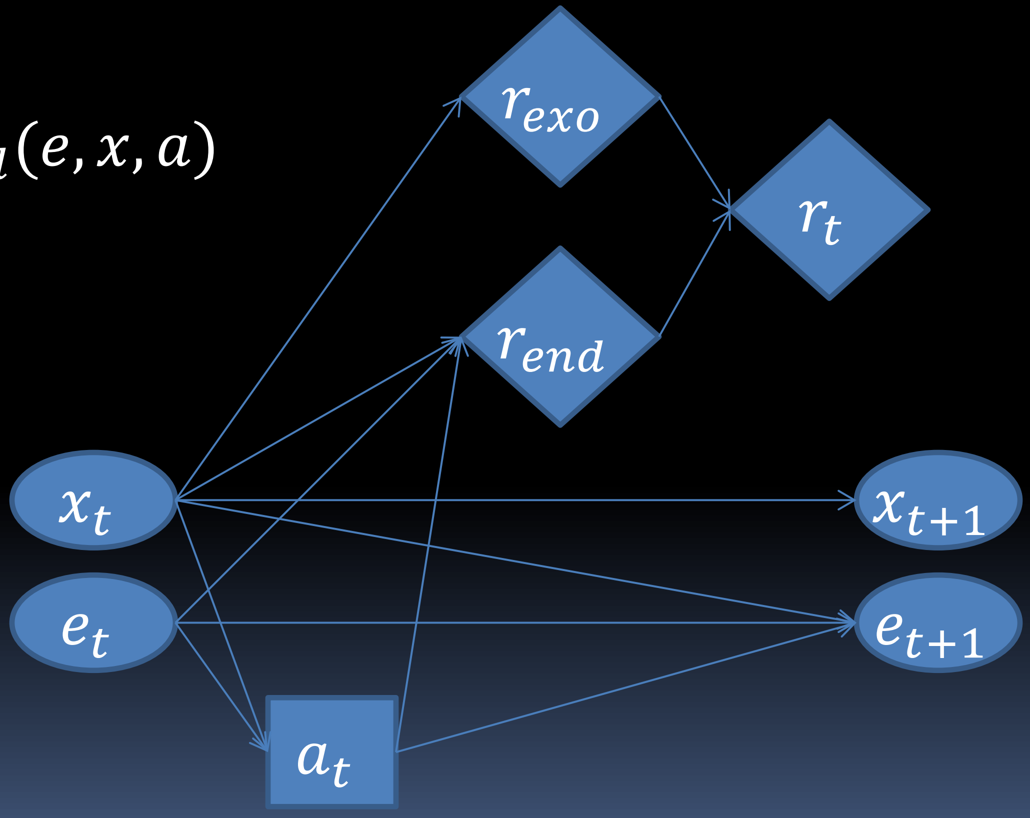
- Transitions:
  - $P(x_{t+1}, e_{t+1} | x_t, e_t, a_t) = P(e_{t+1} | x_t, e_t, a_t) P(x_{t+1} | x_t)$

Actions only affect e_{t+1} (and r_t)
x evolves independently but is still Markov

Approach

- Assumption: Reward Decomposes Additively

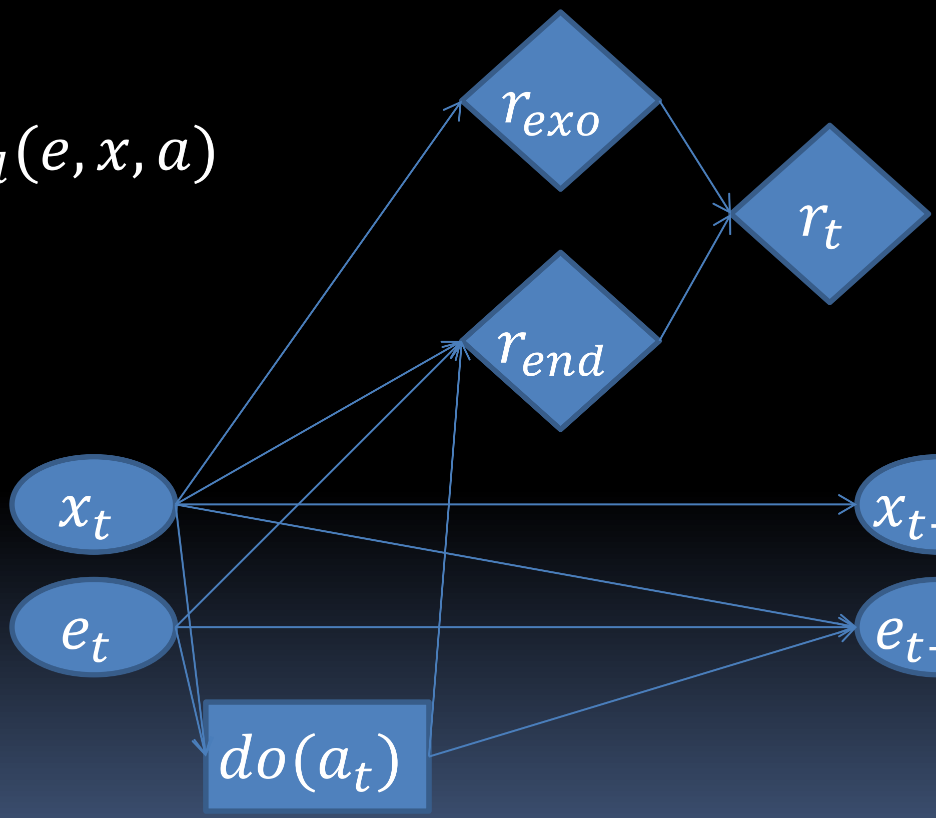
$$R(e, x, a) = R_{exo}(x) + R_{end}(e, x, a)$$



Approach

- Assumption: Reward Decomposes Additively

$$R(e, x, a) = R_{exo}(x) + R_{end}(e, x, a)$$

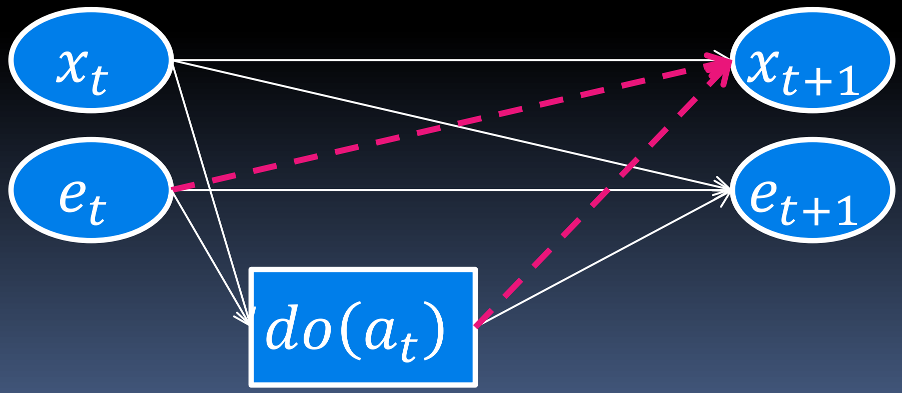


Approach

- Decompose s into (e, x) by enforcing mutual information constraints
  - $(e, x) = F(s)$
  - Solve a regression problem to predict $r_t = R_{exo}(x_t)$
  - How much of the reward can be explained by the exogenous state alone?
- Subtract $r_t - R_{exo}(x_t) = r_{end}$ to obtain the endogenous reward (plus any noise in R_{exo})
- Find an MDP policy π that optimizes just r_{end}

Estimating the Endo-Exo Decomposition

- Suppose we have a database of transitions $\{(s_i, a_i, r_i, s_i')\}_{i=1}^{n}$ gathered by executing one or more exploration policies on the MDP
- Linear case ⇒ additive decomposition:
$$x = W^\top s;\ e = s - WW^\top s$$
- Find W to satisfy $I(x_{t+1}; (e_t, a_t) | x_t) = 0$



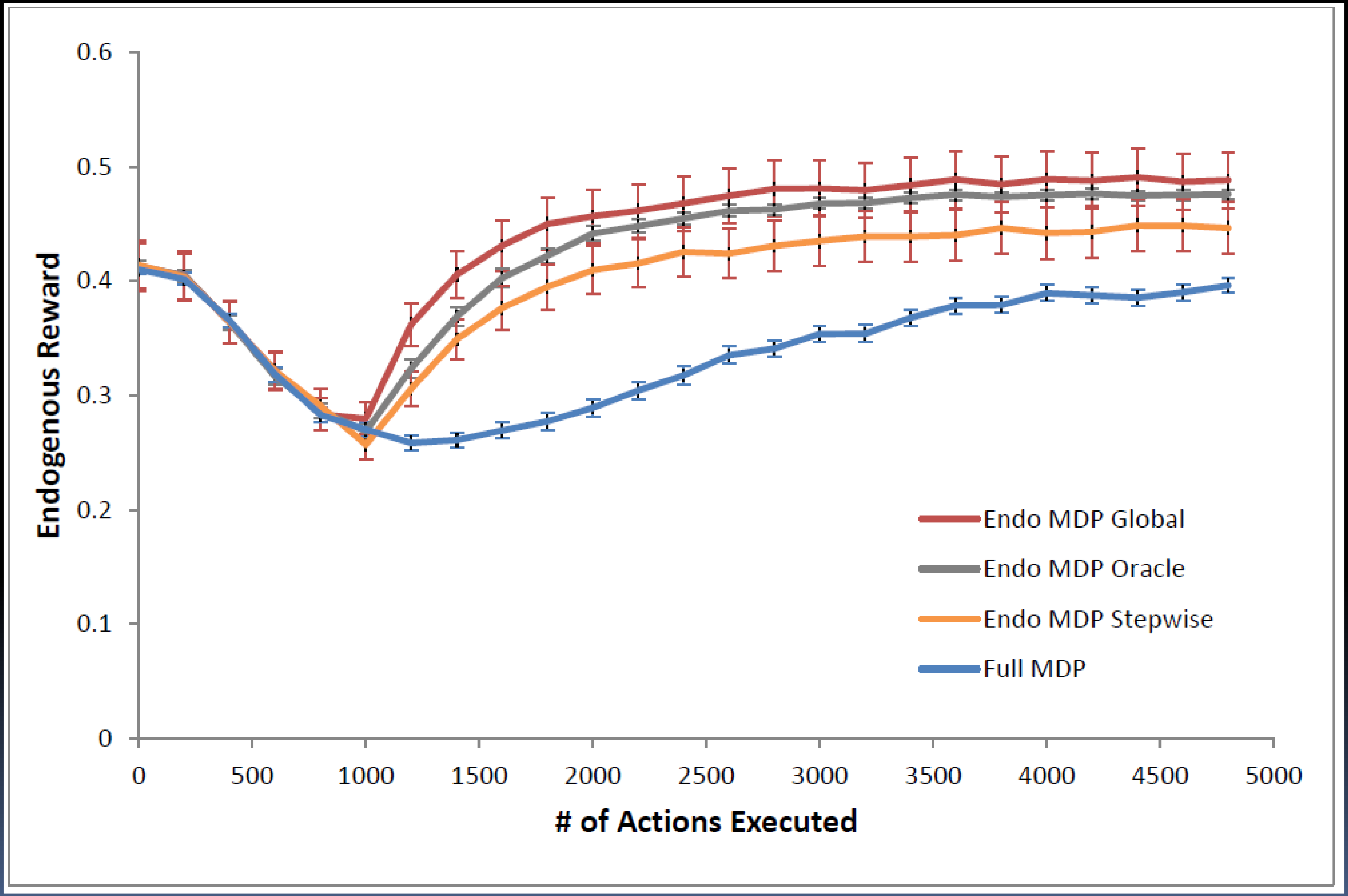
Two Algorithms

- Approximate $I(x_{t+1};(e_t, a_t)|x_t)$ by the Partial Correlation Coefficient
- Global Algorithm
  - For each $1 \leq d_x \leq d$, compute a d-dimensional W
  - Solves d Steiffel Manifold optimizations of increasing size
- Stepwise Algorithm
  - Similar to PCA
  - Compute one column of W in each iteration
  - Solves d 1-dimensional Steiffel Manifold optimizations
- Matlab Manopt

Toy Problem 1: 30 Dimensions

- 15 dimensions are exogenous
- 15 dimensions are endogenous
- $X_{t+1} = M_x X_t + \mathcal{E}_x$
- $E_{t+1} = M_e \begin{bmatrix} E_t \\ X_t \\ A_t \end{bmatrix} + \mathcal{E}_e$
- $\mathcal{E}_x \sim \mathcal{N}(0,0.09)$; $\mathcal{E}_e \sim \mathcal{N}(0,0.04)$
- $S_t = M \begin{bmatrix} E_t \\ X_t \end{bmatrix}$
- $R_x = -3\, avg(X)$; $R_e = \exp[-|avg(E_t) - 1|]$
- M, M_x, M_e are random matrices with elements $\sim\mathcal{N}(0,1)$. Rows normalized to sum to 0.99.
- $\beta = 1$, learning rate = 0.05. 2 hidden layers w/ 40 tanh units

Results



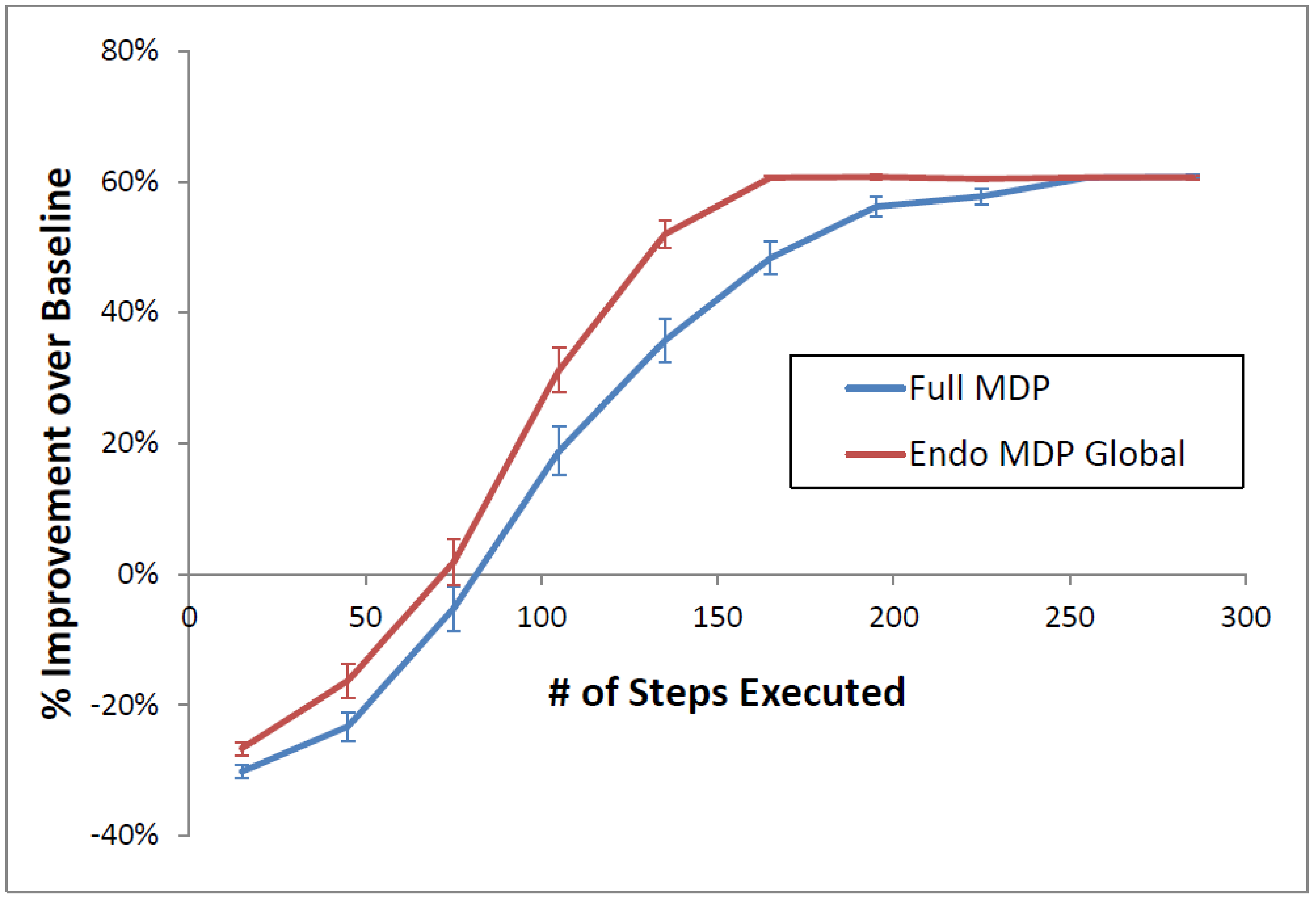
Endogenous Reward
0.6
0.5
0.4
0.3
0.2
0.1
0
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
# of Actions Executed
Endo MDP Global
Endo MDP Oracle
Endo MDP Stepwise
Full MDP

Cell Network Optimization

- Adjust cell tower parameters to minimize # of users experiencing poor throughput
- Action: increase/reduce threshold on signal power for when to switch channel for a user
- Time step: 1 hour
- Data: 5 days, hourly, 105 cells, Huawei Customer
- Simulator: MFMC (Fonteneau et al 2012)
- discount factor 0.95
- features: # active users, avg # of users, channel quality index, small packets/total packets; small packet bytes / total packet bytes

- Reward function: $R_t = -P_t =$ fraction of customers with low bandwidth during period $(t, t + \Delta t)$
- Separate fixed horizon evaluation trials

Results



80%
60%
40%
20%
0%
-20%
-40%
% Improvement over Baseline
0
50
100
150
200
250
300
# of Steps Executed
Full MDP
Endo MDP Global

Summary

- Exogenous state variables can increase reward variance and impede RL
- We can identify these variables by solving an optimization problem with conditional mutual information constraints
- We can then remove the mean effect of the exogenous state

Open Questions & Next Steps

- Identify and Remove Exogenous Noise?
  - Can we also remove the effect of aleatory variation in the exogenous state?
- Irrelevant State Variables
  - We can set up a similar mutual-information problem to identify a subspace that is irrelevant to r_t even though it *is* affected by our actions
- Conditional Causation
  - Is there any benefit to identifying regions of the state space where our actions affect only a portion of e_t?

Reflections on Causal Modeling in Machine Learning

- Confounding is a threat to successful generalization
  - It is one of the key reasons that ML methods do not generalize well
- ML should fit causal models whenever possible

Causal Modeling and Machine Learning

- Pearl (et al.): To make causal inferences, we must make causal assumptions
  - To learn from data, we must make some assumptions (adopt a model space)
  - This can easily be a space of causal models
  - So the causal assumptions can be quite weak

Is fitting causal models different from fitting acausal models?

- Yes
- The “evidence” for fitting a statistical model is just the data
- The “evidence” for fitting a causal model includes the data-generating process
  - randomization, mixing
  - interventions (e.g., instrumental variables)
  - etc.

Is there a unified theory of fitting causal models?

- To fit statistical models, MLE and MAP methods minimize the (penalized) KL divergence between the fitted model and the data
- Can we cast the fitting of causal models into some similar "distance" framework?
  - At present, we have a growing collection of techniques. Can we unify them?

Methodological Benefit of Causal Modeling

- Interventions and Transportability encourage the modeler to explicitly think about how the deployment environment will differ from the training environment
- This is not unique to causal modeling but
  - Causal models provide a vocabulary for expressing many kinds of change
- Prediction: ML will increasingly focus on “threats to generalization” in our search for robustness

Acknowledgments

- Dietterich's time was supported by a gift to OSU from Huawei, Inc.

Questions?