Robust Structured Prediction for Process Data

Xiangli Chen

Computer Science Department
University of Illinois at Chicago

Committee: Brian D.Ziebart (Chair and advisor, UIC CS)
Piotr J.Gmytraisewicz (UIC CS), Tanya Y.Berger-Wolf (UIC CS)
Byron Boots (Georgia Tech IC), Umar Ali Syed (Google)

April 14, 2017

Outline

What is process ?

process

Cambridge Dictionary

noun [C] · **US** /ˈprɑs·es, ˈproʊ·ses/

- **a series of actions or events performed to make something or achieve a particular result, or a series of changes that happen naturally:**

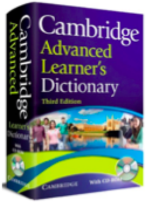
  Completing his degree at night was a long process.

  Graying hair is part of the aging process.

  We are still ***in the process of*** *redecorating the house (= working to decorate it).*

- **A process is also a method of doing or making something, as in industry:**

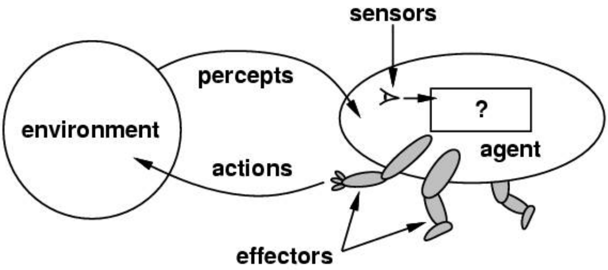
  A new process has been developed for removing asbestos.



Why is it important ?

Problem Formulation-An Interaction

A process often arises within an **interaction** between an **agent** and its **environment**.



Problem Formulation - Concepts

Performing a process:

$$s_1, a_1, \cdots, a_{T-1}, s_T$$

- **State** s contains necessary information
- **Policy** $\pi(a_t|s_t)$
- **Dynamics** $\tau(s_t|a_{1:t-1}, s_{1:t-1})$

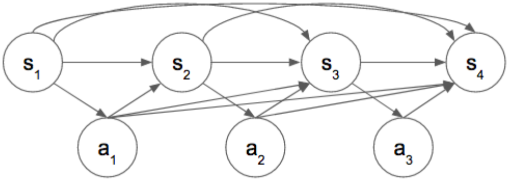


Figure 1: Process

Problem Formulation-Process Prediction

Definition

A **process prediction** problem is an estimation task that given a list of training samples

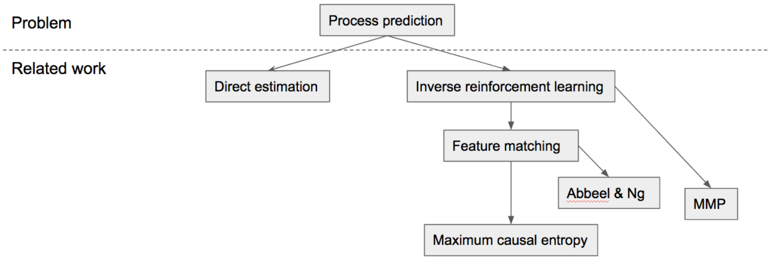
$$\{s_1, a_1, \cdots, s_{T_{n-1}}, a_{T_n-1}, s_{T_n}\}_{n=1}^{N}$$

we want to estimate a policy $\hat{\pi}_t(a_t|s_t)$ with respect to a performance evaluation method.

Evaluation - empirical loss on test samples

- Log loss - how likely
- $0-1$ loss, square or absolute loss

Related Work



Direct Estimation

Direct estimation (Behavioral Cloning) - using classical supervised learning methods (Pomerleau 1989;Sammut+ 1992)

$$p_{H^*}(a|s)$$

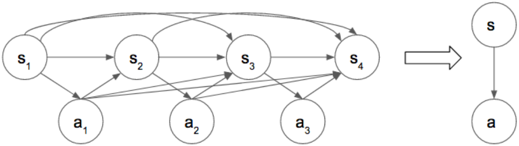
$$H^* = \arg\min_{H \in \mathcal{H}} \text{Empirical training Loss } (H)$$

- Discrete case: classification, e.g. logistic regression, SVM, neural network
- Continuous case: linear regression

Direct Estimation

Direct estimation

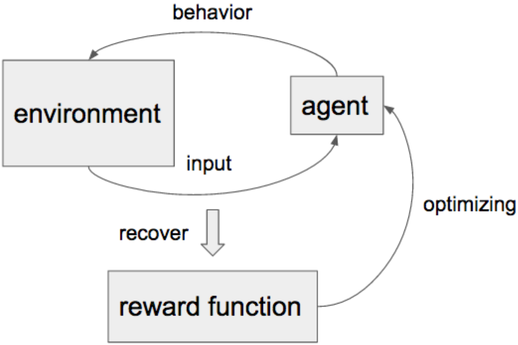
- can't fully experss the structure of the process



- not adaptive - lack of generalization ability
  E.g. driving learning task - driving circumstance changes

Inverse Reinforcement Learning

Inverse reinforcement learning (IRL) (inverse optimal control (IOC))- model learning using **reinforcement learning** (Russell, 1998)



Background

Process is performed to achieve a particular result

Maximizing reward or minimizing cost - succinct, robust and natural description

Minimum jerk principle

Primates limb movement (Hogan, 1984)

Minimum hiring cost

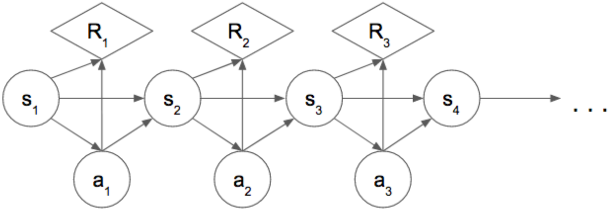
Firm's hiring behavior (Sargent, 1978)

Minimum torque change

Human multijoint arm movement (Uno, 1989)

Infinite-Horizon Markov decision process (MDP)

- dynamics $\tau(s_{t+1}|s_t, a_t)$
- reward function $R(s, a)$ - immediate payoff
- policy $\pi(a|s)$



Value Function and other Decision models

Value function - expected long term reward

$$v_\pi(s) = \mathbb{E}\left[\sum_{k=1}^{\infty} \gamma^{k-1} R_{t+k-1} \middle| s_t = s\right] \quad (\gamma \in (0,1))$$

Optimal policy

$$\hat{\pi} = \arg\max_{\pi} V^{\pi}(s), (\forall s)$$

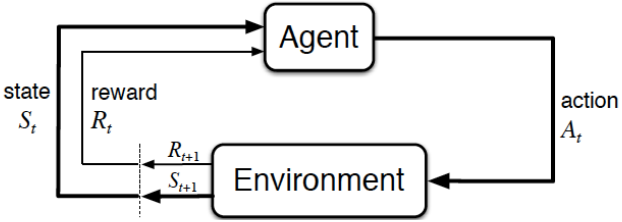
A planning problem - dynamic programming
Finite-Horizon MDP - non stationary (time dependent)
Partially Observable MDP (POMDP) - states unknown

- $\pi(a_t | a_{1:t-1}, o_{1:t})$ - computation intractable
- $\pi(a_t | b_t)$ - belief state b summaries historical information

Reinforcement Learning-Overview

Reinforcement learning (Sutton,Barto,1998)

- Infinite-Horizon MDP
- dynamics is unknown



Learn optimal policy from interaction

Challenge

Specifying reward/cost functions is challenging (Russell, 1998)

- Prior hypothesis may be wrong (horses' gait selection)
- Hard to weight and combine multiattribute reward (running, bee nectar ingestion)

Horses' gait selection - Not for energetic economy

Running - speed, efficiency, stability against perturbations, wear and tear on muscles, etc

Bee nectar ingestion - flight distance, time and risk from wind and predators, etc

Inverse Reinforcement Learning-Principle

Inverse reinforcement learning (IRL) (inverse optimal control(IOC)) (Kalman,1964;Boyd+ 1994;Ng, Russell 2000)

- Assumption - expert is optimal
- Approach - find a reward function R^* that explains the expert's behavior π^*:

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R^*(s_t)|\pi^*\right] \geq \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R^*(s_t)|\pi\right] \;\; \forall \pi$$

Challenging

Find a reward function R^* that explains the expert policy π^*:

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R^*(s_t) | \pi^*\right] \geq \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R^*(s_t) | \pi\right] \quad \forall \pi$$

Challenges:

- Ambiguity - many optimal reward funcitons e.g. $R = 0$
- Complexity - need enumerate all policies
- Infeasibility - imperfect expert policy
- π^* unknown - only expert demonstration

Feature Based Reward Function

Feature based reward function $\Leftrightarrow R(s) = \omega^T \phi(s)$:

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t) | \pi\right] = \omega^T \mu(\pi)$$

Finding R^* equals to finding ω^* such that:

$${\omega^*}^T \mu_E \geq {\omega^*}^T \mu(\pi) \quad \forall \pi$$

Only expert demonstration $\{s_0^i, s_1^i, \cdots\}_{i=1}^m$ in practice:

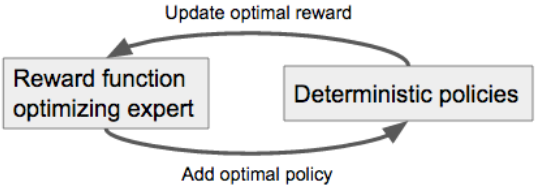
$$\mu_E = \frac{1}{m} \sum_{i=1}^{m} \sum_{t=0}^{\infty} \gamma^t \phi(s_t^i)$$

Feature Matching-Principle

Feature matching (Abbeel, Ng, 2004) - solve the infeasibility
Assume $||\omega||_1 \leq 1$
(optimal policies of $k\omega^T\phi(s)$ are the same for $k > 0$)

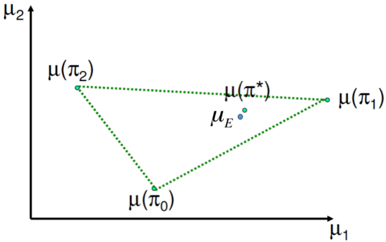
$$||\mu(\pi) - \mu_E||_2 \leq \epsilon \Longrightarrow ||\omega^T\mu(\pi) - \omega^T\mu_E||_2 \leq \epsilon$$

Algorthim



Feature Matching-Issue

Optimal stochstic π^* - mixed by deterministic ones π_0, π_1, π_2



Issue

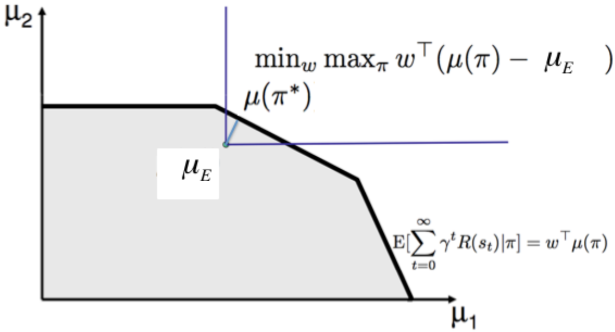
π_0, π_1, π_2 - on the convex hull of π's (extrem points) - performance of π^* has high variance

Min-Max Feature Expectation Matching

Matching imperfect expert may not be appropriate

A Game-Theoretic approach (min-max) (Syed, Schapire. 2008)

Assume $\omega \geq 0, \sum_i \omega_i = 1$ - form a zero-sum game



Maximum Margin Planning

Max margin planning (MMP) (Taskar+ 2005; Ratliff, Bagnell, Zinkevich 2006) - solve the ambiguity

- ${\omega^*}^T \mu_E \geq {\omega^*}^T \mu(\pi) \quad \forall \pi$
- maximize margin

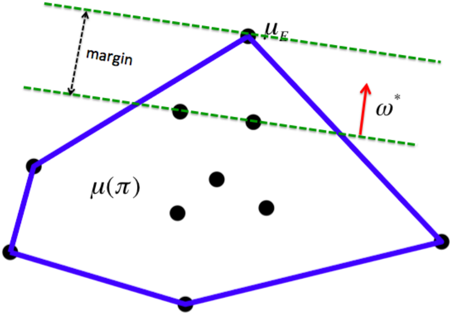


Figure 2: Convex set of π's

MMP-Imperfect Expert Demonstration

Imperfect expert demonstration μ_E - within the interior

- infeasibility exists

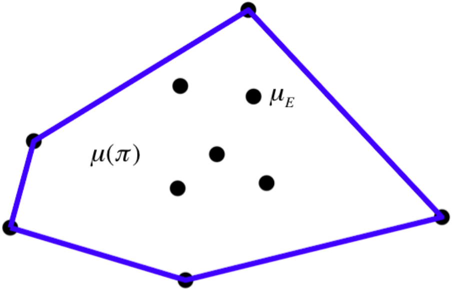
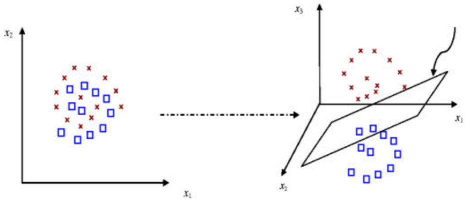


Figure 3: Convex set of π's

MMP-Methods to Imperfect Expert Demonstration

Methods to imperfect expert demonstration

- Project feature vector $\phi(s)$ to high dimensional space



- Soft maximum margin - add slack variables $\xi^{(i)}$'s

$$\min_{\omega,\xi} \parallel \omega \parallel_2^2 + C \sum_i \xi^{(i)}$$

$$\text{s.t. } \omega^T \mu_E^{(i)} \geq \omega^T \mu(\pi^{(i)}) + m(\pi_E^{(i)}, \pi^{(i)}) - \xi^{(i)} \quad \forall i, \pi^{(i)}$$

MMP-Issues

Maximum margin planning

$$\min_{\omega,\xi} \| \omega \|_2^2 + C \sum_i \xi^{(i)}$$

$$\text{s.t. } \omega^T \mu_E^{(i)} \geq \omega^T \mu(\pi^{(i)}) + m(\pi_E^{(i)}, \pi^{(i)}) - \xi^{(i)} \quad \forall i, \pi^{(i)}$$

Issues

- Feature projection leads to poor generalization
- Tradeoff between maximum margin and slackness
- Very large number of constraints
- Sensitive to imperfect expert demonstration

Robust Structure

Model interaction process via **Maximum Causal Entropy** (Ziebart,Bagnell,Dey 2010)

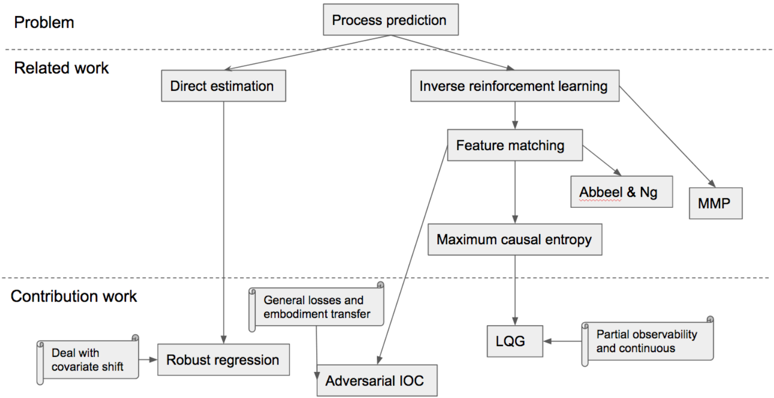
$$p(s_{1:T}, a_{1:T}) = \underbrace{\prod_{t=1}^{T} p(s_t | s_{1:t-1}, a_{1:t-1})}_{\text{provided process}} \times \underbrace{\prod_{t=1}^{T} p(a_t | a_{1:t-1}, s_{1:t})}_{\text{unknown process}}$$

Optimal policy $\{p_t(a_t|a_{1:t-1}, s_{1:t})\}_{t=1}^{T}$:

$$\max \underbrace{\mathbb{E}_p \left[-\sum_{t=1}^{T} \log p(a_t | a_{1:t-1}, s_{1:t}) \right]}_{\sum_{t=1}^{T} H(a_t|a_{1:t-1}, s_{1:t})}$$

Under $\mathbb{E}_p [\mathcal{F}(A_{1:T}, S_{1:T})] = \tilde{\mathbb{E}}_p [\mathcal{F}(A_{1:T}, S_{1:T})]$

Contribution Work



Research Contribution

Develop **robust** structure prediction models for **process** data that

- allows partially observable continuous environments (Chen,Ziebart 2015)
- deals with covariate shift (Chen,Monfort,Liu,Ziebart 2016)
- enables various imitation learning evaluation measures and embodiment transfer (Chen,Carr,Ziebart 2015; Chen,Monfort,Ziebart,Carr 2016)

Principle of Robustness

What does **robustness** mean ?

Principle of Robustness

What does **robustness** mean ?

- As uncertain as possible (Jaynes 1957)
- Best estimation under the worst case (Topϕse 1979; Gr$\ddot{u}$nwald+ 2004)

Maximum Uncertainty

H measures uncertainty of a random event p requires

- continuity, monotonic increasing and consistency

The only H is known as **Shannon Entropy** (Shannon 1948)

$$H = -K \sum_{i=1}^{n} p_i \log p_i \quad (K \text{ is just a constant})$$

Estimate p

$$\max_{p \in \Delta} H(p) \ (\Delta \text{ represents constraint})$$

Maximum entropy - as uncertain as possible - prevent bias

Optimal Encoding

Optimal prefix-free encoding of sending messages a, b, c, d:

distribution	1/2	1/4	1/8	1/8
prefix-free code	0	10	110	111

Length of optimal prefix-free encoding is close to

$$-\log p(x)$$

Minimum expected prefix-free encoding length

$$\mathbb{E}_p[-\log p(X)] \leq \underbrace{\mathbb{E}_p[-\log q(X)]}_{\text{expected log loss}}$$

Expected log loss measures the "distance" of p and q

Min-Max Estimation Approach

Best estimation under the worst case

$$q^* = \arg\min_{q \in \Xi} \max_{p \in \Gamma} \mathbb{E}_p[-\log q(X)]$$

Choosing q^* achieves the minimum loss bound

$$\mathbb{E}_{p\in\Gamma}[-\log q^*(x)] \leq \min_{q \in \Xi} \max_{p \in \Gamma} \mathbb{E}_p[-\log q(X)] \leq \max_{p \in \Gamma} \mathbb{E}_p[-\log q_{\in\Xi}(X)]$$

Strong duality holds (if Γ is closed and convex)

$$\max_{p \in \Gamma} H(p) = \max_{p \in \Gamma} \min_{q \in \Xi} \mathbb{E}_p[-\log q(X)] = \min_{q \in \Xi} \max_{p \in \Xi} \mathbb{E}_p[-\log q(X)]$$

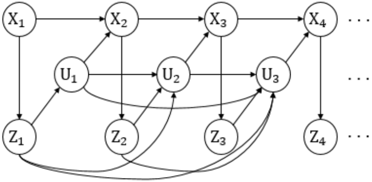
and

$$p^* = \arg\max_{p \in \Gamma} H(p) = q^*$$

Problem Formulation

Linear qudratic Gaussian system

(X : state; U : action; Z : observation)



$$X_1 \sim N(\mu, \Sigma_{d_1}) \quad X_{t+1}|x_t, u_t \sim N(Ax_t + Bu_t, \Sigma_d)$$
$$Z_t|x_t \sim N(Cx_t, \Sigma_o)$$

Need to obtain the control policy $f(u_t|u_{1:t-1}, z_{1:t})$

Optimal Control Apporach

Optimal control - minimize the expected cost (**cost matrix** M):

$$\min \mathbb{E}\left[\sum_{t=1}^{T+1} X_t^T M X_t\right]$$

Optimal control policy (closed form):

$$u_t = -L_t \hat{x}_t(+) \quad \hat{x}_t(+) = \mathbb{E}[X_t | \zeta_t]$$

- ζ_t : sufficient statistics of $z_{1:t}, u_{1:t-1}$
- L_t : feedback gain - recursively defined

Optimal Control Apporach

What if the cost matrix M is not given ?

Our approach - learn the policy from training samples

Maximum Uncertainty Approach

$$f(x_{1:T}, z_{1:T}, u_{1:T}) = \underbrace{\prod_{t=1}^{T} f(x_t, z_t | x_{1:t-1}, z_{1:t-1}, u_{1:t-1})}_{\text{provided process}} \times \underbrace{\prod_{t=1}^{T} f(u_t | u_{1:t-1}, x_{1:t}, z_{1:t})}_{\text{unknown process}}$$

Causal conditional probability

$$f(u_{1:T} || x_{1:T}, z_{1:T}) = \prod_{t=1}^{T} f(u_t | u_{1:t-1}, x_{1:t}, z_{1:t})$$

The **causal entropy** measures the uncertainty over the interaction process.

$$H(U_{1:t} || X_{1:T}, Z_{1:T}) = \mathbb{E}[-\log f(U_{1:T} || X_{1:T}, Z_{1:T})] = \sum_{t=1}^{T} H(U_t | X_{1:t}, Z_{1:t}, U_{1:t-1})$$

Principle of Maximum Causal Entropy

$H(U_{1:T}||X_{1:T}, Z_{1:T})$

- nonconvex function of $f(u_t|u_{1:t-1}, x_{1:t}, z_{1:t})_{t=1}^{T}$
- convex function of $f(u_{1:T}||x_{1:T}, z_{1:T})$

Causal Simplex of LQG

Causal simplex Ξ of $f(u_{1:T}||x_{1:T}, z_{1:T})$ (Chen Ziebart 2015):

$f(u_{1:T}||x_{1:T}, z_{1:T}) \in \Xi$

is equivalent to

$$f(u_{1:T}||x_{1:T}, z_{1:T}) = \prod_{t=1}^{T} f(u_t|u_{1:t-1}, x_{1:t}, z_{1:t})$$

Ξ is a convex set of $f(u_{1:T}||x_{1:T}, z_{1:T})$
Formulate a convex optimization problem (Γ is convex)

$$\max_{f \in \Gamma} H(U_{1:T}||X_{1:T}, Z_{1:T}) \; (\Xi \subseteq \Gamma)$$

Min-Max Interpretation

Minimum expected prefix-free encoding length over the interaction process

$$\mathbb{E}_f\left[-\sum_{t=1}^{T}\log f(U_t|U_{1:t-1},X_{1:t},Z_{1:t})\right] = \mathbb{E}_f\left[-\log f(U_{1:T}||X_{1:T},Z_{1:T})\right]$$
$$\leq \underbrace{\mathbb{E}_f\left[-\log h(U_{1:T}||X_{1:T},Z_{1:T})\right]}_{\text{expected causal log loss}}$$

Strong duality holds (Γ is closed and convex)

$$\min_{h\in\Xi}\max_{f\in\Gamma}\mathbb{E}_h[-\log f] = \max_{f\in\Gamma}\min_{h\in\Xi}\mathbb{E}_h[-\log f] = \max_{f\in\Gamma} H(U_{1:T}||X_{1:T},Z_{1:T})$$

Maximum Causal Entropy Inverse LQG

Definition

The **maximum causal entropy inverse LQG** (convex optimization problem)

$$\max_{\{f(u_{1:T}||z_{1:T},x_{1:T})\}\in\Xi} H(U_{1:T}||Z_{1:T},X_{1:T})$$

$$\text{such that } \mathbb{E}_f\left[\sum_{t=1}^{T+1} X_t X_t^T\right] = \tilde{\mathbb{E}}\left[\sum_{t=1}^{T+1} X_t X_t^T\right]$$

Motivation of expectation constraint (Abbeel, Ng 2004)

$$\mathbb{E}_f = \tilde{\mathbb{E}} \implies \mathbb{E}_{f,M} = \tilde{\mathbb{E}}_M \quad (\forall M \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|})$$

Gaussian Belief Solution

Theorem

Belief state b_t summaries the historical information $u_{1:t-1}, z_{1:t}$

$$X_t | b_t \sim N(\mu_{b_t}, \Sigma_{b_t})$$

Optimal control policy:

$$U_t | \mu_{b_t} \sim N\left(-W_t \mu_{b_t}, \Sigma_{U_t}\right)$$

Relates to LQG Optimal Control

Optimal policy - optimal control law:

$$u_t = -L_t \hat{x}_t(+) \quad \hat{x}_t(+) = \mathbb{E}[X_t | \zeta_t]$$

Optimal policy - our min-max approach:

$$U_t | \mu_{b_t} \sim N\left(-W_t \mu_{b_t}, \Sigma_{U_t}\right)$$

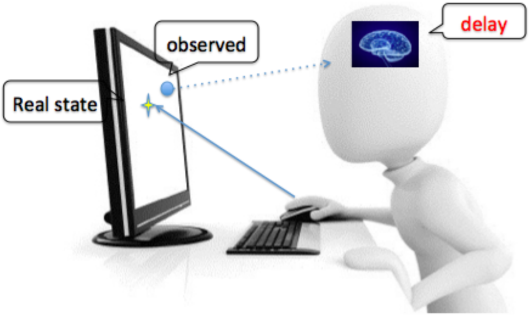
Theorem

Given the Lagrangian multiplier matrix M as the cost matrix:

$$W_t = L_t$$

Real Experiment-Problem

Modeling mouse cursor pointing motions



Real Experiment-Data

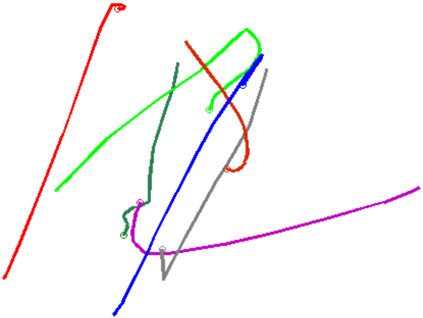


Figure 4: Example mouse cursor trajectories terminating at small circle positions exhibiting characteristics of delayed feedback.

Real Experiment-Direct Estimation

Direct estimation k^{th}-order Markov model ($k = 1, 2, 3, 4$)
(linear regression model)

$$\hat{\vec{s}}_t = [\vec{s}_{t-1} \ \vec{s}_{t-2} \ \ldots \ \vec{s}_{t-k}]\vec{\alpha} + \epsilon \qquad \epsilon \sim N(0, \sigma^2)$$

- $\vec{s}_t \triangleq [x_t \ y_t]$ position of mouse cursor
- control policy $\hat{\vec{u}}_t = \hat{\vec{s}}_t - \vec{s}_{t-1}$

Real Experiment-Linear Quadratic

Linear Quadratic setting

$$\vec{x}_t \triangleq [x_t\ y_t\ \dot{x}_t\ \dot{y}_t\ \ddot{x}_t\ \ddot{y}_t]^T \quad \vec{u}_t = [\dot{x}_t\ \dot{y}_t]^T$$

$$\begin{pmatrix} \dot{x}_t \\ \dot{y}_t \end{pmatrix} = \begin{pmatrix} x_t - x_{t-1} \\ y_t - y_{t-1} \end{pmatrix} \quad \begin{pmatrix} \ddot{x}_t \\ \ddot{y}_t \end{pmatrix} = \begin{pmatrix} \dot{x}_t - \dot{x}_{t-1} \\ \dot{y}_t - \dot{y}_{t-1} \end{pmatrix}$$

- Linear Quadratic Regulator (LQR) - fully observable
- LQG - delay - partially observable

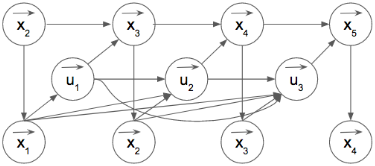


Figure 5: Example LQG setting with one time step delay

Real Experiment-Result

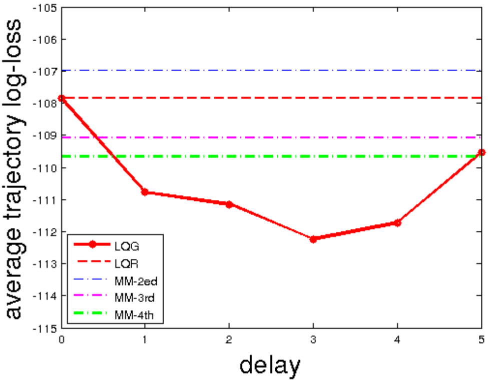


Figure 6: Average trajectory log-loss of: the LQG model with various amounts of delay, t_0; the LQR model; Markov models of order 2,3,4.

Research Contribution

Develop **robust** structure prediction models for **process** data that

- allows partially observable continuous environments (Chen,Ziebart 2015)
- deals with covariate shift (Chen,Monfort,Liu,Ziebart 2016)
- enable various imitation learning evaluation measures and embodiment transfer (Chen,Carr,Ziebart 2015; Chen,Monfort,Ziebart,Carr 2016)

Motivation

What if the training samples are not representative ?

Non-representative

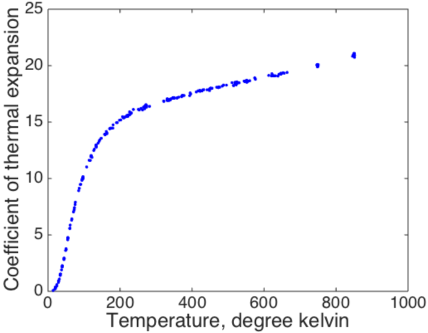


Figure 7: Hahn1 dataset representing the result of a National Institute of Standards and Technology (NIST) study of the thermal expansion of copper.

Background

A **loss function** $L(x, y, f(x;\theta))$ measures the discrepancy
$f(x;\theta)$ is the **approximation function** to y
Let

$$f^*(x) = \mathbb{E}_{Y|X=x}[Y|X = x]$$

A model is correctly specified if there exists a θ^* such that

$$f(x;\theta^*) = f^*(x).$$

Otherwise, the model is misspecified.

Optimal estimator

Optimal parameter

$$\theta^* = \arg\min_\theta [L(X, Y, f(X;\theta))].$$

Optimal estimator

$$\hat{\theta} = \arg\min_\theta \left[\frac{1}{N} \sum_{i=1}^{N} L(x_i, y_i, f(x_i;\theta)) \right] \left(\lim_{N_{tr}\to\infty} \hat{\theta}_{tr} \xrightarrow{p} \theta^*_{tr} \right)$$

If training p_{tr} and test p_{te} share the same distribution

$$p_{tr}(x, y) = p_{te}(x, y)$$

$\hat{\theta}_{tr}$ is a consistent estimator of θ^*_{te} even for misspecified models:

$$\lim_{N_{tr}\to\infty} \hat{\theta}_{tr} \xrightarrow{p} \theta^*_{tr} = \theta^*_{te}.$$

Covariate Shift

In many real-world applications,

$$p_{tr}(x, y) \neq p_{te}(x, y)$$

Covariate shift

- $p_{tr}(x) \neq p_{te}(x)$
- $p_{tr}(y|x) = p_{te}(y|x)$

Under covariate shift $\hat{\theta}_{tr}$ is

- consistent estimator of θ^*_{te} if the model is correctly specified
- no longer consistent if the model is misspecified

Importance weighted Method

Importance weighted method (Shimodaira, 2000) requires importance ratio's first moment to be finite (Cortes+, 2010)

$$\mathbb{E}_{p_{tr}(x)}\left[p_{te}(X)/p_{tr}(X)\right] < \infty.$$

Reweight loss function

$$\mathbb{E}_{X^{te},Y^{te}}\left[L(X,Y,f(X;\theta))\right] = \mathbb{E}_{X^{tr},Y}\left[\frac{p_{te}(X)}{p_{tr}(X)}L(X,Y,f(X;\theta))\right].$$

So that

$$\theta^*_{te} = \theta^*_{triw} = \arg\min_{\theta} \mathbb{E}_{X^{tr},Y}\left[\frac{p_{te}(X)}{p_{tr}(X)}L(X,Y,f(X;\theta))\right].$$

Hence $\hat{\theta}_{triw}$ is a consistent estimator of θ^*_{te} that

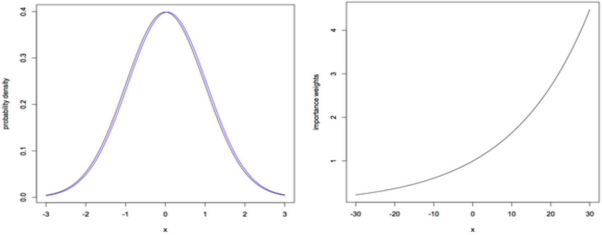
$$\lim_{N_{tr}\to\infty} \hat{\theta}_{triw} \xrightarrow{p} \theta^*_{triw} = \theta^*_{te}$$

Importance-Weighted Method - Issue

Too restricted to hold

$$\mathbb{E}_{p_{tr}(x)}\left[p_{te}(X)/p_{tr}(X)\right] < \infty$$

(e.g., two Gaussian distributions with slightly shifted mean)



Linear regression

Ordinary least squares (linear regression)

$$\hat{y}_{a,b}(x) = b^T x + a$$

(Adaptive) **Importance Weighted Regression** (Sugiyama+, 2012)

$$\operatorname*{argmax}_{a,b,\sigma} \mathbb{E}_{\tilde{f}_{\text{tr}}(x)\tilde{f}(y|x)} \left[\left(\frac{f_{\text{te}}(X)}{f_{\text{tr}}(X)} \right)^{\gamma} \log \hat{f}_{a,b,\sigma}(Y|X) \right]$$

where $\gamma \in (0,1)$ is the flattening parameter.
Ordinary least squares ($\gamma = 0$), and importance weighted ($\gamma = 1$) at its extrems.

Robust Bias-Aware Regression

Definition

The **robust bias-aware regression estimator**, $\hat{f}(y|x)$, ($f_o(Y|X)$ works as a baseline distribution):

$$\min_{\hat{f}(y|x)} \max_{f(y|x)\in\Xi} \mathbb{E}_{f_{te}(x)f(y|x)}\left[-\log\frac{\hat{f}(Y|X)}{f_o(Y|X)}\right]$$

$$\text{such that: } \mathbb{E}_{f_{\text{tr}}(x)f(y|x)}[\Phi(X,Y)] = \frac{1}{n}\sum_{i=1}^{n}\phi(x_i,y_i)$$

Mean and Variance

If $f_o(y|x) = N(\mu_o, \sigma_o^2)$ and $\Phi(x,y) = [y\ x^T\ 1]^T[y\ x^T\ 1]$
Lagrangian multiplier matrix:

$$M = \begin{bmatrix} M_{(y,y)} & M_{(y,x1)} \\ M_{(x1,y)} & M_{(1,1)} \end{bmatrix}$$

The robust bias-aware regression:

$$\hat{f}_M(y|x) \sim N(\mu(x,M), \sigma^2(x,M))$$

$$\mu(x,M) = \left(2\frac{f_{tr}(x)}{f_{te}(x)}M_{(y,y)} + \frac{1}{\sigma_o^2}\right)^{-1}\left(-2\frac{f_{tr}(x)}{f_{te}(x)}M_{(y,x1)}\begin{bmatrix} x \\ 1 \end{bmatrix} + \frac{1}{\sigma_o^2}\mu_o\right)$$

$$\sigma^2(x,M) = \left(2\frac{f_{tr}(x)}{f_{te}(x)}M_{(y,y)} + \frac{1}{\sigma_o^2}\right)^{-1}$$

The distribution's certainty is moderated by $f_{tr}(x)/f_{te}(x)$.

Base Distribution

$f_o(y|x) = N(\mu_o, {\sigma_o}^2)$ is a Gaussian distribution with mean and variance estimated from the range

$$[y_{\text{min}}, y_{\text{max}}]$$

of y's of the source dataset D_{src}:

$$\mu_o = \frac{y_{\text{min}} + y_{\text{max}}}{2}, \quad \sigma_o^2 = \left(\frac{y_{\text{max}} - \mu_o}{2}\right)^2.$$

Hence all of the y's of the source dataset are located within the 95% confidence of the base distribution.

Comparison Approaches

- RBA_{KLD} - our approach (relative log loss)
- BS - baseline Gaussian distribution
- RBA_{DE} - robust-biase aware regression via differential entropy (log loss)
- LS - ordinary least squares $Y|X \sim N(b^T X + a, \sigma^2)$
- BAIWLS - best adaptive importance weighted least squares (optimal flattening parameter $\gamma \in \{0.1, 0.2, \cdots, 0.9\}$)
- IWLS - importance weighted least squares
- BLR - Bayesian linear regression (prior $[b^T a] \sim N(0, I)$)

Datasets

Regression datasets from the UCI repository

Table 1: Datasets for empirical evaluation

Dataset	#Examples	#Features	Output
Airfoil	1503	5	sound pressure
Concrete	1030	8	strength
Housing	506	14	value of home
Music	1059	66	latitude
Crime	1994	127	crime rate
Parkinsons	5725	16	UPDRS score
WineQuality	6497	11	quality score
IndoorLocation	21048	529	latitude

Constructing Datasets with Bias

Consider both synthetically created and naturally occuring bias

Table 2: Experimental settings

Dataset	#Source	#Target	Bias Setting
Airfoil	150-751	752	synthetic
Concrete	100-515	515	synthetic
Housing	75-253	253	synthetic
Music	160-529	530	synthetic
Parkinsons	1430-2862	2863	synthetic
Crime	40-278	1716-1954	different state
WineQuality	4898	1599	different color
Parkinsons	1877	1839	different age
IndoorLocation	9371	10566	different floor

Experimental Result

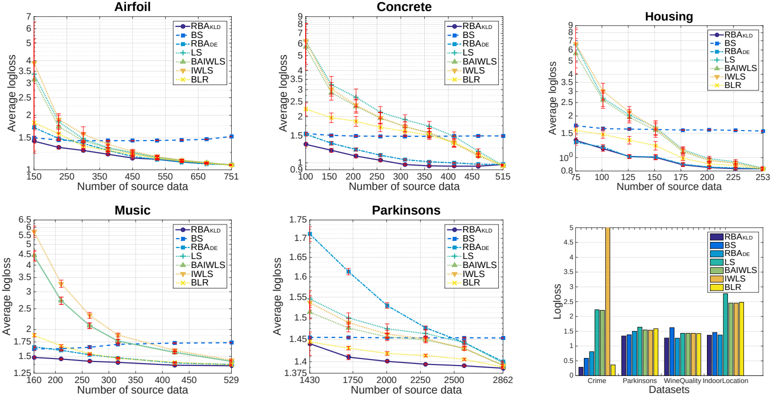


Figure 8: Five plots of the average empirical logloss for target datasets with 95% confidence interval. A bar figure showing empirical log loss on four natural bias datasets.

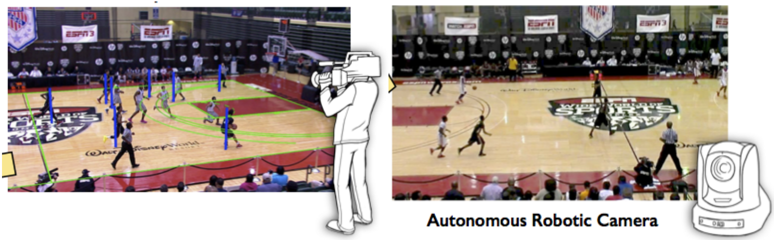
Research Contribution

Develop **robust** structure prediction models for **process** data that

- allows partially observable continuous environments (Chen,Ziebart 2015)
- deals with covariate shift (Chen,Monfort,Liu,Ziebart 2016)
- enables various imitation learning evaluation measures and embodiment transfer (Chen,Carr,Ziebart 2015; Chen,Monfort,Ziebart,Carr 2016)

Adversarial IOC-Motivation

A basketball match



Want the robotic camera mimic the operator

- **Square or absolute** loss are much preferable
- Robotic camera is **less capable** than an operator

Adversarial IOC-Motivation

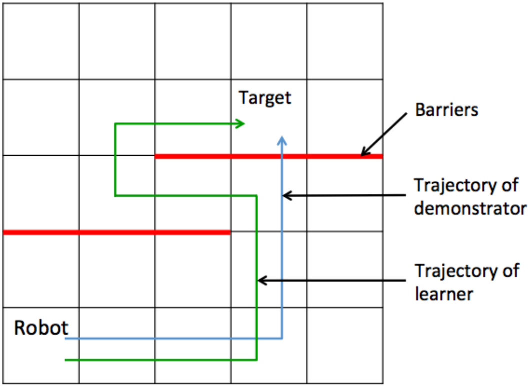


Figure 9: Grid world navigation

Problem Definition

Definition

In the task of **imitation learning with general losses and embodiments**,

- Demonstrated samples from a distribution under
  - a known dynamics $\tau(s_{1:T}||a_{1:T})$
  - an unknown control policy $\pi(a_{1:T}||s_{1:T})$
- Learner attempts to choose
  - a control policy $\hat{\pi}(\hat{a}_{1:T}||\hat{s}_{1:T})$
  - for potentially different dynamics $\hat{\tau}(\hat{s}_{1:T}||\hat{a}_{1:T})$
- Minimize a general loss function

$$\min_{\hat{\pi}} \text{loss}_{\tau,\hat{\tau}}(\pi, \hat{\pi})$$

Adversarial Approach

Definition

The **adversarial inverse optimal control learner** is defined as a **zero-sum game**:

$$\min_{\hat{\pi}} \max_{\check{\pi} \in \tilde{\Xi}} \mathbb{E}\left[\sum_{t=1}^{T} \text{loss}(\hat{S}_t, \check{S}_t) \middle| \check{\pi}, \tau, \hat{\pi}, \hat{\tau}\right]$$

$\tilde{\Xi}$ represents constraints measured from demonstrated data

$$\check{\pi} \in \tilde{\Xi} \iff \mathbb{E}[\sum_{t=1}^{T} \phi(\check{S}_t)|\check{\pi}, \tau] = \tilde{c} \triangleq \mathbb{E}[\sum_{t=1}^{T} \phi(S_t)|\tilde{\pi}, \tilde{\tau}]$$

Loss functions that additively decompose over the state sequence - computation benefit

Definition of Fisher Consistency

Definition

An imitation learning algorithm producing policy π_{imit} is **Fisher consistent** if, given

- the demonstrator's control policy π for any demonstrator/imitator decision processes, $(\tau, \hat{\tau})$
- a sufficiently expressive feature representation for policies

the policy π_{imit} is a loss minimizer:

$$\pi_{\text{imit}} \in \operatorname*{argmin}_{\hat{\pi}} \mathbb{E}\left[\text{loss}_{\tau,\hat{\tau}}(\pi, \hat{\pi})\right].$$

Fisher Consistency of Adversarial Approach

Theorem

Given a sufficiently rich feature representation defining the constraint set Ξ, the adversarial inverse optimal control learner is a **Fisher consistent loss function minimizer** *for all additive, state-based losses.*

Proof

A sufficiently rich feature representation is equivalent to the constraint set Ξ containing only the true policy π. Then, under $\check{\pi} = \pi$, then reduces to:

$$\min_{\hat{\pi}} \mathbb{E}\left[\sum_{t=1}^{T} \text{loss}(\hat{S}_t, \check{S}_t) \middle| \pi, \tau, \hat{\pi}, \hat{\tau}\right]$$

which is the loss function minimizer.

Generalization Bound

Theorem

The adversarial formulation provides a generalization bound:

$$P(\pi \in \tilde{\Xi}) \geq 1 - \alpha \implies P\left(\mathbb{E}\left[\sum_{t=1}^{T} loss(\hat{S}_t, S_t) | \pi, \tau, \hat{\pi}, \hat{\tau}\right] \geq \mathbb{E}\left[\sum_{t=1}^{T} loss(\hat{S}_t, \check{S}_t) | \check{\pi}, \tau, \hat{\pi}, \hat{\tau}\right]\right) \leq \alpha.$$

Proof

If $\pi \in \tilde{\Xi}$, then:

$$\mathbb{E}\left[\sum_{t=1}^{T} \text{loss}(\hat{S}_t, S_t) | \pi, \tau, \hat{\pi}, \hat{\tau}\right] \leq \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \text{loss}(\hat{S}_t, \check{S}_t) | \check{\pi}, \tau, \hat{\pi}, \hat{\tau}\right]}_{\text{min max bound}}$$

Dual Form

Theorem

An **equilibrium** *for the game is obtained by solving an unconstrained zero-sum game parameterized by a vector of Lagrange multipliers* ω*:*

$$\min_{\omega} \underbrace{\min_{\hat{\pi}} \max_{\check{\pi}} \mathbb{E}\left[\sum_{t=1}^{T} loss(\check{S}_t, \hat{S}_t) + \omega \cdot \phi(\check{S}_t) \middle| \check{\pi}, \tau, \hat{\pi}, \hat{\tau}\right]}_{\textit{zero-sum game}} - \omega \cdot \tilde{c}.$$

Using gradient descent method (λ learning rate)

$$\omega \leftarrow \omega - \lambda \cdot (\mathbb{E}_{P(\check{S}_{1:T}, \check{A}_{1:T})}[\sum_{t=1}^{T} \phi(\check{S}_t) | \check{\pi}^*, \tau] - \tilde{c})$$

Payoff Matrix

Stochastic policy of each player (demonstrator $\check{\pi}$ or learner $\hat{\pi}$) is a mixture of **deterministic policies**: $\check{\delta}$ and $\hat{\delta}$.

	$\check{\delta}_1$	$\check{\delta}_2$	$\dots$	$\check{\delta}_k$
$\hat{\delta}_1$	$\ell(\check{\delta}_1, \hat{\delta}_1)$ $+\psi(\check{\delta}_1)$	$\ell(\check{\delta}_2, \hat{\delta}_1)$ $+\psi(\check{\delta}_2)$	$\dots$	$\ell(\check{\delta}_k, \hat{\delta}_1)$ $+\psi(\check{\delta}_k)$
$\hat{\delta}_2$	$\ell(\check{\delta}_1, \hat{\delta}_2)$ $+\psi(\check{\delta}_1)$	$\ell(\check{\delta}_2, \hat{\delta}_2)$ $+\psi(\check{\delta}_2)$	$\dots$	$\ell(\check{\delta}_k, \hat{\delta}_2)$ $+\psi(\check{\delta}_k)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\hat{\delta}_j$	$\ell(\check{\delta}_1, \hat{\delta}_j)$ $+\psi(\check{\delta}_1)$	$\ell(\check{\delta}_2, \hat{\delta}_j)$ $+\psi(\check{\delta}_2)$	$\dots$	$\ell(\check{\delta}_k, \hat{\delta}_j)$ $+\psi(\check{\delta}_k)$

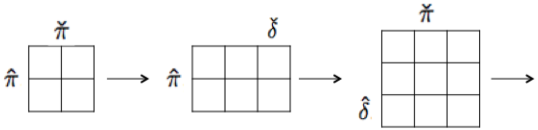
Table 3: The payoff matrix with $\ell(\check{\delta}, \hat{\delta}) = \mathbb{E}[\sum_{t=1}^{T} \text{loss}(\check{S}_t, \hat{S}_t) | \check{\delta}, \tau, \hat{\delta}, \hat{\tau}]$ and $\psi(\check{\delta}) = \omega \cdot \mathbb{E}[\sum_{t=1}^{T} \phi(\check{S}_t) | \check{\delta}, \tau]$.

Double Oracle Method

The payoff matrix grows exponentially

Double oracle method (McMahan, Gordon, Blum, 2003)

- Solve sub-game via linear programming
- Add best responses
- Repeat until convergence



Best response - solve a Finite-Horizon MDP

$$\operatorname*{argmin}_{\hat{\delta}} \mathbb{E}_{\check{\pi}}; \ \text{or} \operatorname*{argmax}_{\check{\delta}} \mathbb{E}_{\hat{\pi}}$$

Synthetic Experiment Setting

Synthetic experiment - navigation across a grid world

- Cost: $C(s) = \theta^T \phi(s) + \varepsilon(s)$
- Transition dynamics:
$$p(s_{t+1} | s_t, a_t) = \begin{cases} p_m & \text{matching the action} \\ \frac{1-p_m}{\#\text{ of neighbor cells}} & \text{neighbor cells} \end{cases}$$
- Loss: euclidean distance between demonstrator and learner

Compare to maximum margin planning (MMP)

Synthetic Experiment Result

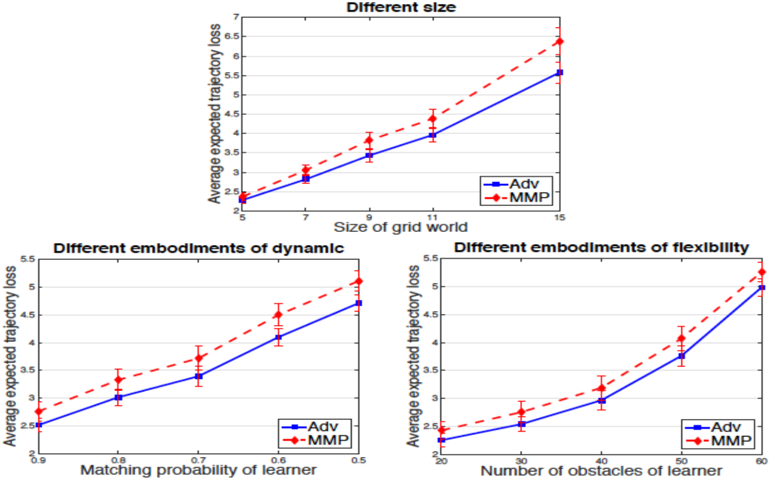


Figure 10: Experimental results with 95% confidence interval of various settings of the grid world's characteristics

Real Experiment Setting

Real experiment-Learning camera control from demonstration

- Output: camera's horizontal pan angle θ_t
- Input: 14-element vector describing players' location X_t
- State: pan angle and its velocity-map them to 305 possible states
- Features: 32 element vector $[\theta, \theta^2, \dot{\theta}, \dot{\theta}^2, \theta X, \dot{\theta} X]$ that is $\phi(s_t)$.

Real Experiment Setting

Comparison methods:

- Linear regression (LS)
- Constrained by camera empirical dynamic (LS_C)
- Condition on previous state (LS_{MI})
- Maximum marginal planning (MMP)
- With start location provided (MMP_{SL})
- Adversarial approach (Adv)
- Wwith start location provided (Adv_{SL})

Real Experiment Result

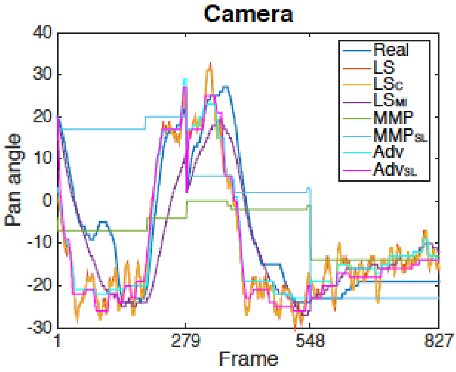


Figure 11: Imitating human camera operator's pan angle control using a regression approach, maximum margin planning, and our adversarial inverse optimal control method.

Real Experiment Result

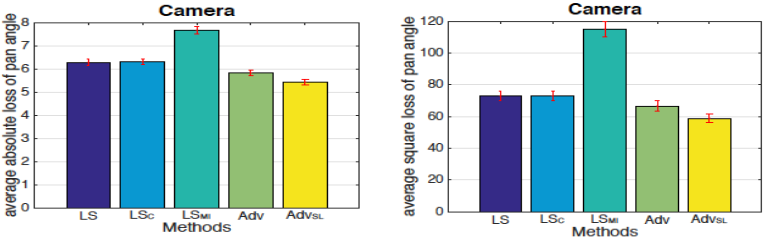


Figure 12: Average squared loss and absolute loss of the imitator (with 95% mean confidence intervals estimates) with maximum margin planning results suppressed due to being significantly worse and off of the presented scale.

Conclusion

Process environment are inherently complex and noisy.

conclusion

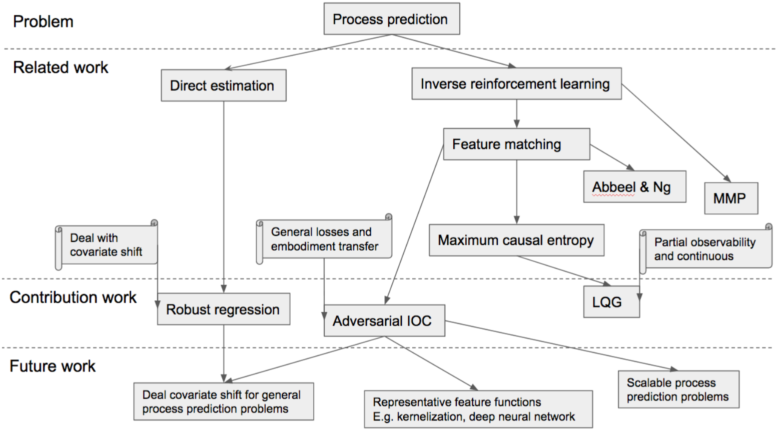
Process prediction models that are developed

- by deriving the best estimation under the worst case
- subject to matching known properties of the real distribution

demonstrate robust prediction performance and benefit from

- incorporating partial observability
- dealing with non-stationary settings
- enabling various evaluation measures and embodiment transfer

Future Work


Problem
Process prediction
Related work
Direct estimation
Inverse reinforcement learning
Feature matching
Abbeel & Ng
MMP
Deal with covariate shift
General losses and embodiment transfer
Maximum causal entropy
Partial observability and continuous
Contribution work
Robust regression
Adversarial IOC
LQG
Future work
Deal covariate shift for general process prediction problems
Representative feature functions
E.g. kernelization, deep neural network
Scalable process prediction problems

Thank You

Many thanks to

- **Advisor** - Pro.Brian D. Ziebart
- **Committee members**
- **Collaborators** - Anqi Liu, Mathew Monfort and Peter Carr (Disney)
- **Lab mates** (chat, discussion, fun, resource and many others) - Kaiser Asif, Rizal Fathony, Sima Behpour, Jia Li, Hong Wang, Wei Xing, Chris Schultz, Sanket Gaurav, Andrea Tirinzoni