A Minimax Robust Approach for Learning to Assist Users with Pointing Tasks

Sima Behpour
Department of Computer Science
University of Illinois at Chicago
Chicago, IL 60607
sbehpo2@uic.edu

Brian D. Ziebart
Department of Computer Science
University of Illinois at Chicago
Chicago, IL 60607
bziebart@uic.edu

Motivations and Challenges

Learning to provide appropriate assistance to people in different situations is an extremely important, but insufficiently investigated machine learning task. Applications include human-robot and human-computer interactions settings (Lieberman 2009; Goodrich and Schultz 2007) to maximizing the benefits of assistive technologies (Hoey et al. 2005). Three key challenges must be overcome to appropriately address this task:

- **Complexity:** the space of possible assistive policies can be very large, making many existing methods (e.g., from reinforcement learning) too data inefficient to be practical.
- **Noise and misspecification:** observed human behavior is often noisy and parametric formulations that reduce complexity will typically suffer from model misspecification, leading to unboundedly sub-optimal assistance.
- **Biasedness:** data available for learning a model is biased by previously provided assistive actions, violating the typical assumptions of supervised learning.

We develop a general framework for learning to assist in single intervention settings. The framework narrows the search for effective assistance by viewing previous behavior under assistance through a restricted set of statistics. Assistive policies for the worst-case context-assistance-outcome relationships satisfying these statistics are obtained. We embed the problem of learning how to assist users in cursor-based target pointing tasks into this framework and outline its usage.

Approach

There are two key intertwined problems underlying the *learning to assist* task (with variables defined and related to one another according to Figure 1): estimating the outcome of assistive actions in a particular context, $P(y|w, x, z)$; and choosing the optimal intervention policy, $P(z|x)$, given the observed context.

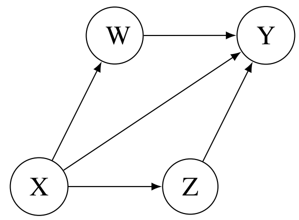

Figure 1: Bayesian network structure relating observed context (X), the unobserved context (W), the intervention (Z) and the task outcome (Y).

Motivated by robust estimation (Grünwald and Dawid 2004) and control (Başar and Bernhard 2008) techniques, we formulate the problem of assisting as a two-player game between an action selecting player and an outcome estimation player (Definition 1).

Definition 1. *The* **minimax assistance game** *is a zero-sum game between: (1) the player who chooses the assistive strategy $P(z|x)$ to minimize the expected task performance loss ($Loss(w, x, y, z)$); and (2) an adversary who chooses a conditional outcome distribution $P(y|x, z)$ for evaluation that maximizes the expected task performance loss[1]:*

$$\min_{\hat{P}(z|x)\in\Delta} \max_{\check{P}(y|w,x,z)\in\Xi\cap\Delta} \mathbb{E}_{\substack{P(x)P(z|x)\\ P(w|x)\check{P}(y|w,\,x,\,z)}} [Loss(W, X, Y, Z)]. \tag{1}$$

We consider feature expectations constraints that define the set Ξ: $\mathbb{E}_{\tilde{P}(x,w,z)\check{P}(y|w,x,z)}[\mathbf{f}(W, X, Y, Z)] = \mathbb{E}_{\tilde{P}(x,w,y,z)}[\mathbf{f}(W, X, Y, Z)]$. We then focus on the dual Lagrangian of the optimization:

$$\begin{aligned}\min_{\theta} \max_{\check{P}(y|w,x,z)} \min_{z(x)} & \mathbb{E}_{P(w,x,y)}[\text{Loss}(W, X, Y, z(X))] \quad (2)\\ & - \theta \cdot \Big(\mathbb{E}_{\tilde{P}(x,w,z)\check{P}(y|w,x,z)}[\mathbf{f}(W, X, Y, Z)] \\ & \quad - \mathbb{E}_{\tilde{P}(x,w,y,z)}[\mathbf{f}(W, X, Y, Z)]\Big).\end{aligned}$$

Intuitively, given the Lagrange multipliers θ, the adversary chooses outcome distribution $\check{P}(y|w, x, z)$ making the best assistive policy $z(x)$ as unattractive as possible while paying a penalty for being different from observed distribution statistics. The appropriate balance between these competing objectives is established by the outer optimization of θ. Additionally, the choice of the feature vector $\mathbf{f}(w, x, y, z)$ naturally limits the complexity of the learning task to be significantly smaller than a typical reinforcement learning or bandit learning formulation.

[1]We use Δ to represent the probability simplex.

Cursor Pointing Assistance Application

We consider the human-computer interaction task of selecting a target object in a computer interface. Providing appropriate assistance to users with motor control impairments for pointing at intended targets would be extremely useful in this setting to improve the efficiency of users' interactions with computers (Balakrishnan 2004; Ziebart, Dey, and Bagnell 2012). We developed an application that presents a sequence of target selection tasks to a user and learns to intervene to improve the efficiency of target selection (Figure 2). The learner is allowed to intervene by choosing the cursor's initial position, but does not know which of the target objects the user must select. The locations and sizes of the targets are chosen randomly, but constrained not to overlap.

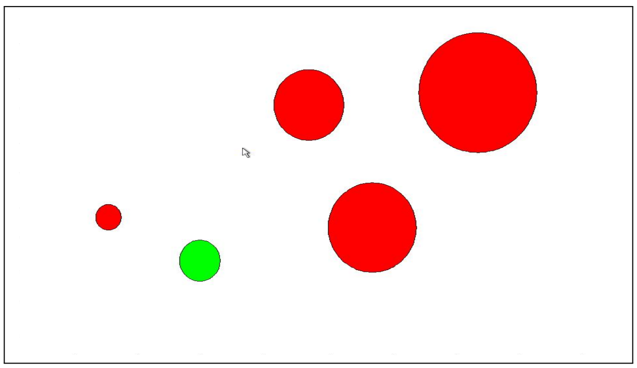

Figure 2: An example of a target pointing task with an intended target (green) and additional targets (red). Target positions and sizes vary between examples.

As an illustrative example, using feature functions of the form $\mathbf{f}(x, w, y, z) = \{y^2, y \log(\frac{\text{distance(w,z)}}{\text{size(w)}} + 1), y\}$ in our framework produces a completion time estimation function[2]:

$$y_\theta(x, w, z) = \frac{\overbrace{-\theta_1 \log\left(\frac{\text{distance(w,z)}}{\text{size(w)}} + 1\right) - \theta_2}^{\text{Fitts's law estimate}}}{2\theta_0}$$
$$+ \begin{cases} \overbrace{\dfrac{P(x)P(w|x)}{2\tilde{P}(x)\tilde{P}(w|x)\tilde{P}(z|x)\;\theta_0}}^{\text{adversarial penalty}} & \text{if z optimal} \\ 0 & \text{otherwise.} \end{cases}$$

The first term of this function matches with Fitts's law estimates of pointing task completion time (Fitts 1954; **?**) with θ parameters estimated from data.

Expected completion times to reach the unknown intended target using this estimate are shown in Figure 3. As the function is non-convex and multi-modal, local gradient-based search methods are employed to find optima (Figure 4). We also plan to investigate using a graphics processing unit to efficiently search for optima exhaustively.

[2]We use distance(w, z) to represent the distance from the intervention point z to the actual target w and size(w) to represent the diameter of target w.

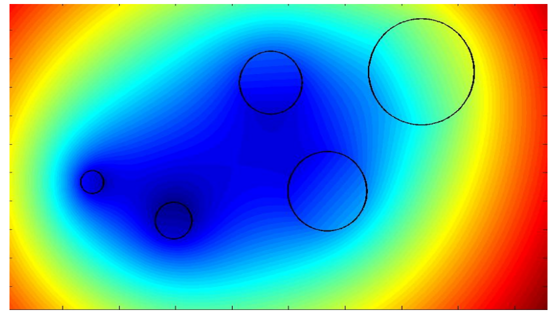

Figure 3: Expected completion times at each possible intervention positions.

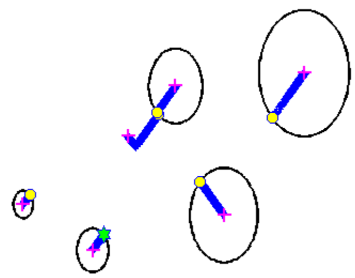

Figure 4: Local optimization of the expected completion time function each reaching a local optima (yellow circle) and a global optima (green star).

References

Balakrishnan, R. 2004. "Beating" Fitts law: virtual enhancements for pointing facilitation. *International Journal of Human-Computer Studies* 61(6):857–874.

Başar, T., and Bernhard, P. 2008. *H-infinity optimal control and related minimax design problems: a dynamic game approach*. Springer.

Fitts, P. M. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47(6):381.

Goodrich, M. A., and Schultz, A. C. 2007. Human-robot interaction: a survey. *Foundations and trends in human-computer interaction* 1(3):203–275.

Grünwald, P. D., and Dawid, A. P. 2004. Game theory, maximum entropy, minimum discrepancy, and robust Bayesian decision theory. *Annals of Statistics* 32:1367–1433.

Hoey, J.; Poupart, P.; Boutilier, C.; and Mihailidis, A. 2005. Pomdp models for assistive technology. In *Proc. AAAI Fall Symposium on Caring Machines: AI in Eldercare*.

Lieberman, H. 2009. User interface goals, ai opportunities. *AI Magazine* 30(4):16.

Ziebart, B. D.; Dey, A. K.; and Bagnell, J. A. 2012. Probabilistic pointing target prediction via inverse optimal control. In *Proceedings of the ACM International Conference on Intelligent User Interfaces*, 1–10.