Inverse Reinforcement Learning from Summary Data

Antti Kangasrääsiö
Samuel Kaski
first.last@aalto.fi

Published in *Machine Learning* (2018) vol 107, pp. 1517-1535

Background

RL models are commonly used for modeling human decision-making. A key problem in this respect is parameter inference based on realistic observation data.

It is common that observation data is incomplete, contains noise, or observations are missing. Occasionally only summarized aggregate observations might be available. However, traditional IRL methods tend to assume full observability.

When small amounts of observation noise is present [1], or few observations are missing [2], EM-type solutions exist for estimating the latent observations, allowing traditional IRL methods to be used. However, this approach is not feasible with significant observation noise or most of the observations missing.

[1] Activity forecasting, Kitani et al. 2012
[2] EM for IRL with hidden data, Bogert et al. 2016

Our Contributions

We demonstrate that parameter inference is possible for RL models even in the presence of arbitrary trajectory-level observation noise, σ(ξ), thus significantly extending the types of situations where RL models can be applied.

We derive the exact Bayesian solution for this problem, but demonstrate that it is very expensive to evaluate. We propose two approximations: a Monte-Carlo estimate and an ABC estimate, which are significantly faster to evaluate.

We demonstrate that the methods allow full posterior inference for a realistic model of human cognition, based on realistic but very restricted observation data.

Take-home message: Regarding partial observability in IRL, there now exists formulations for three different situations:
(1) Agent has partial observability of the environment state → POMDP model
(2) External observer has partial observability on *state* level
→ traditional IRL methods can be extended
New: (3) External observer has partial observability on *trajectory* level
→ presented methods can be used

IRL-SD Problem

Assume an agent is behaving optimally within an MDP environment, producing trajectories $\{\xi_1, \xi_2, \ldots, \xi_N\}$. Further assume an observation noise function $\sigma(\xi) = P(\xi_\sigma|\xi)$ which hides the true trajectories from the external observer, and that the MDP is not fully known to the external observer.

The IRL from Summary Data (IRL-SD) problem is then:

Given observations $\{\xi_{\sigma 1}, \xi_{\sigma 2}, \ldots, \xi_{\sigma N}\}$, the function σ and a prior $P(\theta)$

Estimate the unknown parameters θ of the MDP

Exact Likelihood

The exact likelihood for the IRL-SD problem is tractable, but very expensive to evaluate as we need to integrate over all plausible true trajectories.

$$L(\theta|\Xi_\sigma) = \prod_{i=1}^{N} P(\xi_{i\sigma}|\theta) = \prod_{i=1}^{N} \sum_{\xi_i \in \Xi_{ap}} P(\xi_{i\sigma}|\xi_i) P(\xi_i|\theta)$$

$$P(\xi_i|\theta) = P(s_0^i) \prod_{t=0}^{T_i-1} \pi_\theta^*(s_t^i, a_t^i) P(s_{t+1}^i|s_t^i, a_t^i)$$

Monte-Carlo Approximation

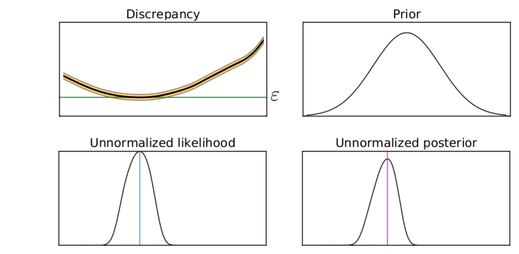
This approach estimates the likelihood based on sampled trajectories. The approximation can be computed as long as we know σ as a distribution $P(\xi_\sigma|\xi)$. However, the approach suffers from numerical problems with rare observations, for which $P(\xi_\sigma|\xi)$ may be zero for all sampled trajectories.

$$\hat{L}(\theta|\Xi_\sigma) = \prod_{i=1}^{N} \frac{1}{N_{MC}} \sum_{\xi_n \in \Xi_{MC}} \frac{P(\xi_{i\sigma}|\xi_n) P(\xi_n|\theta)}{P(\xi_n|\theta)}$$

$$= \prod_{i=1}^{N} \frac{1}{N_{MC}} \sum_{\xi_n \in \Xi_{MC}} P(\xi_{i\sigma}|\xi_n).$$

ABC Approximation

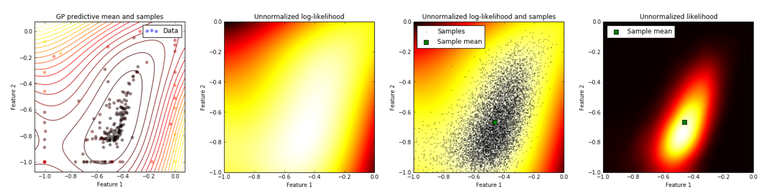
The ABC approximation uses the discrepancy between the Monte Carlo sample and the observation data for estimating the likelihood. Thus it only requires that can be evaluated, and it does not suffer from the same numerical problems as MC. The discrepancy function needs to be chosen; often the prediction error function is suitable.



Discrepancy: $\delta(\Xi_\sigma^A, \Xi_\sigma^B) \to [0, \infty)$
$d_\theta \sim \delta(\Xi_\sigma^{sim}, \Xi_\sigma)$

Likelihood: $\tilde{L}_\varepsilon(\theta|\Xi_\sigma) = P(d_\theta \leq \varepsilon|\theta)$

Inference

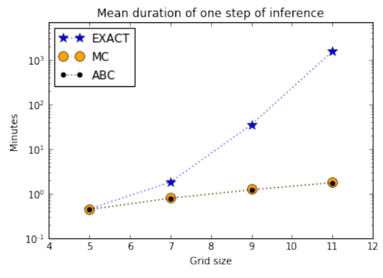


We estimated the likelihood-surfaces from samples using Gaussian Processes and Bayesian Optimization. We then drew samples using MCMC to estimate the shape and mean of the distribution.

Experiments (simulated data)

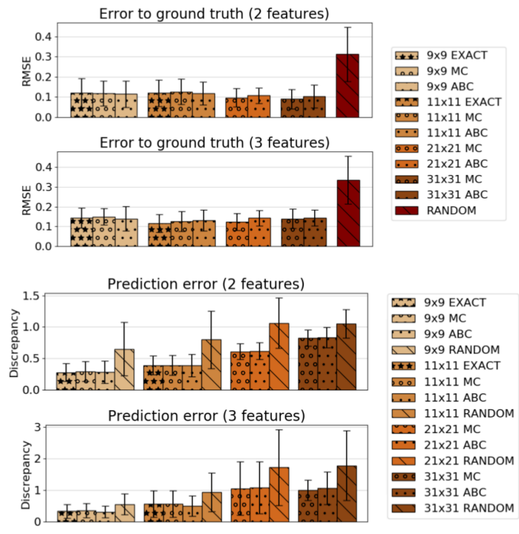
We performed experiments with grid world environments of various sizes and with variable number of unknown parameters.

Algorithm runtime (one step) with different grid sizes

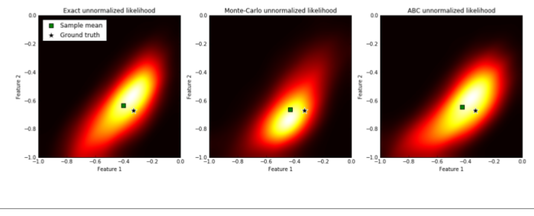


We demonstrate that the approximate methods scale significantly better than exact solution, while maintaining similar inference performance.

Inference quality (error to ground truth, prediction error) with different grid sizes and dimensionality of reward function



Inferred likelihood densities

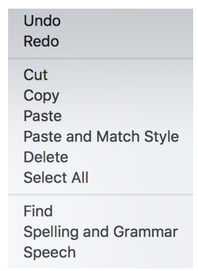


Experiments (real data)

The task of the user was to repeatedly search for a given item from a drop-down menu, and click the item if present. The unknown parameters were fixation duration, click delay and probability of recalling menu layout from memory.

The observation only contained the duration of the episode, and whether the user found the item or not (ie. no states observed, only final action known)

Inference was possible with ABC, resulting in good model fit and informative posterior.

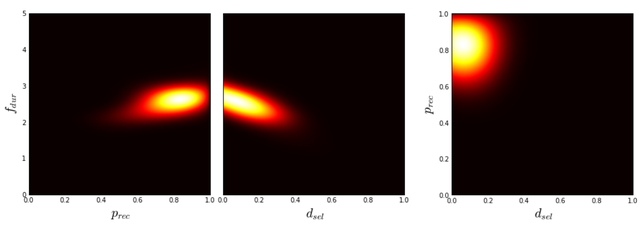


Predictions vs. observations (model fit)

Feature	MAP	Observation data
TCT (abs)	430 ms	470 ms
TCT (pre)	980 ms	970 ms
Saccades (abs)	1.4	1.9
Saccades (pre)	3.1	2.2

The condition when the target is absent from the menu is denoted by (abs), and the condition when the target is present by (pre)

Approximate posterior distribution



Posterior demonstrates that fixation duration was well identified, while more uncertainty remains of the the other two parameters.

See also earlier work: Kangasrääsiö *et al.* **Inferring Cognitive Models from Data using Approximate Bayesian Computation.**
In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI'17, pp. 1295-1306. ACM, 2017.