Inferring actions, intentions, and causal relations in a neural network

Andrew M. Saxe
Harvard University

Overview

Even from a young age, we can select actions to achieve desired goals, infer the goals of other agents, and learn causal relations in our environment through social interactions. Crucially, these abilities are productive or generative: we can impute desires to others that we have never held ourselves.

This capacity has been captured by the powerful Bayesian Theory of Mind formalism (Baker, Saxe, & Tenenbaum, 2011), but it remains to forge connections to the rich neural data around action selection, goal inference, and social causal learning. How can productive inference about actions and intentions arise within the neural circuitry of the brain?

Using the recently-developed linearly solvable Markov decision process, we present a neural network model which permits a distributed representation of tasks. Such a representation allows the expression of infinite possibilities by combining a finite set of bases, enabling truly generative inference of actions, goals, and causal relations in a neural network framework.

The key added ingredient is intentional actions: the neural network is constrained by its architecture to interpret data as arising from efficient action selection towards a goal.

Distributed tasks with the LSMDP

To represent novel tasks in a neural network, a natural approach is a distributed representation, in which new tasks are expressed as a blend of previously learned tasks.

Optimal policies in standard MDP formulations, however, do not compose.

The Linearly Solvable Markov Decision Process (LSMDP) is an alternative MDP formulation in which tasks do blend optimally (Todorov, 2006; Todorov, 2009; Saxe, 2015).

Instantaneous reward: $\mathcal{R}(s,a) = \underbrace{R(s)}_{\text{Seek reward}} - \lambda \underbrace{\mathrm{KL}\left(a(\cdot|s)||P(\cdot|s)\right)}_{\text{Act efficiently}}$

Optimal action: $a^*(s'|s) = \dfrac{P(s'|s)Z(s')}{\mathcal{G}[Z](s)}$

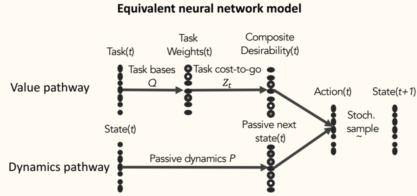
Probability of trajectory through states $s_1, s_2, \cdots, s_N$ (Dvijotham & Todorov, 2010):

$$L[Z,P] = -\sum_n \log Z(s_{n+1}) + \sum_n \log \sum_s P(s|s_n)Z(s) + \sum_n \log P(s_{n+1}|s_n)$$

Bellman equation: $(I - Q_i P_i)z_i = Q_i P_b z_b \quad Z = e^V$ = desirability function

Distributed task representation: $q = Q_t w \implies Z = Z_t w$

Equivalent neural network model

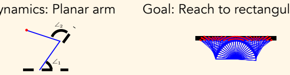


Choosing Actions

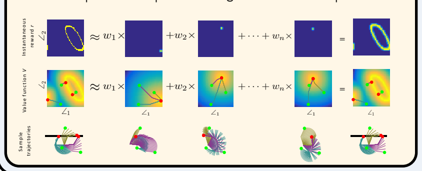
Goals → Actions ← Dynamics

Reaching in space

Dynamics: Planar arm Goal: Reach to rectangular region



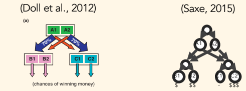
Task decomposition: Express as weighted sum of component tasks

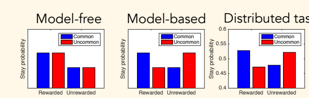


Outcome revaluation in sequential choice

- Simple two stage choice task with varying terminal rewards
- Adaptation to new rewards thought to require model-based control
- Distributed task model behaves like highly flexible model-based agent for tasks in the span of the distributed representation

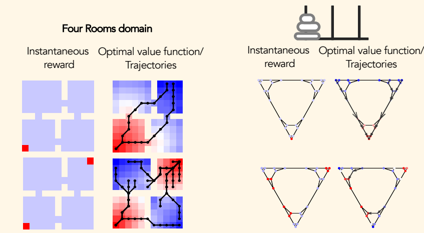
Standard formulation (Doll et al., 2012) LMDP formulation (Saxe, 2015)





Latent learning in spatial navigation

After random exploration of a maze environment, introduction of a reward at one location leads to instant goal-directed behavior towards that point (Tolman, 1948)

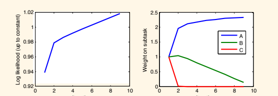


Inferring Intentions

Actions → Goals ← Dynamics

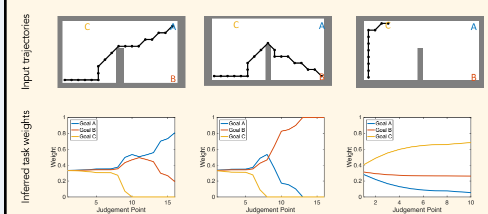
- Inferring goals is the inverse reinforcement learning problem
- Typically expensive: requires solving RL in the inner loop
- For LMDPs, just iterate (Dvijotham & Todorov, 2010):

$$\Delta w = \frac{\partial L}{\partial w} = \sum_n \underbrace{\left(\frac{1}{w^T Z_t^T s_{n+1}} Z_t^T s_{n+1} - \frac{1}{w^T Z_t^T P^T s_n} Z_t^T P^T s_n \right)}_{\text{Upweight tasks that value this transition highly}}$$



Spatial goal inference

- Humans can infer likely goals from partial trajectories in spatial environments (Baker, Saxe, & Tenenbaum, 2009)
- Distributed task representations enable similar inferences in a neural network framework

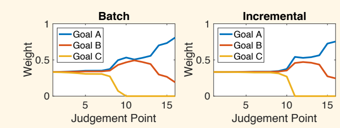


Online inference

- Exact ML inference in this model requires computing the gradient over the entire trajectory after each update of the iterative inference, which requires memory of the sequence
- Can derive an approximate online version by applying update after each transition:

$$\Delta w = -\alpha w + \frac{1}{w^T Z_t^T s_{n+1}} Z_t^T s_{n+1} - \frac{1}{w^T Z_t^T P^T s_n} Z_t^T P^T s_n$$

- This scheme is fully incremental and online, and could explain how inferences about commonly-experienced goals can become almost effortless and habitual



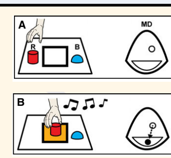
Inferring Causal Relations

Goals → Dynamics ← Actions

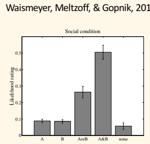
- Observational causal learning: Inferring causal structure from observations of social actions and goals (Meltzoff, Waismeyer, & Gopnik, 2012; Goodman, Baker & Tenenbaum, 2009)
- Update passive dynamics matrix *P*, which encodes the causal structure of which states follow each other
- Iterative procedure based on subgradient descent on *P*:

$$\Delta P = \frac{\partial L}{\partial P}$$

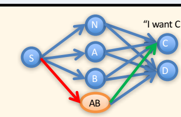
- After each update, project back to probability simplex
- Implicitly incorporates the principle of efficiency (Csibra, 2013), i.e., that actions are taken to efficiently achieve their ends

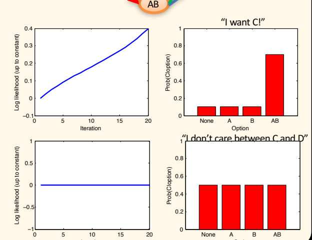


Waismeyer, Meltzoff, & Gopnik, 2014



Goodman, Baker, & Tenenbaum, 2009





Conclusions

Distributed task representations via the LMDP allow cheap inference of actions, goals, and causal relations

The neural network implementation is architecturally constituted to reason about intentional actions

The assumption of *efficient* actions is embedded in the cost function, and enables inference of goals and causal structure from social observation (Csibra, 2013)

The model requires only incremental online updates, and may provide a starting point for investigating neural circuits capable of complex goal inference

Causality is operationalized as the knowledge needed to compute an optimal action to achieve unknown rewards in the future (Woodward, 2003; Gopnik & Schulz, 2007).

Many limitations remain:
- Compositionality only holds at boundary absorbing states of the first exit MDP
- Modern Bayesian Theory of Mind models reason about structured goal sequences (Nakahashi et al., 2016) and beliefs (Baker et al., 2011)

Funding: Swartz Foundation