Learning to Distinguish Between Belief and Truth

Stefano V. Albrecht

June 15, 2016

Department of Computer Science
The University of Texas at Austin

Structure

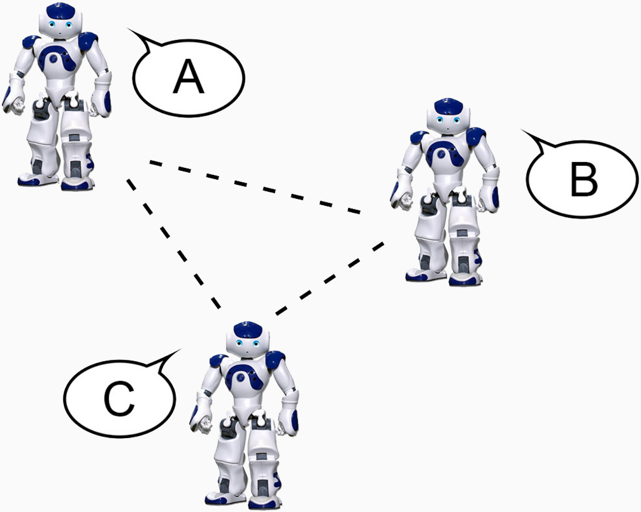
Introduction

Behavioural Hypothesis Testing

The Future

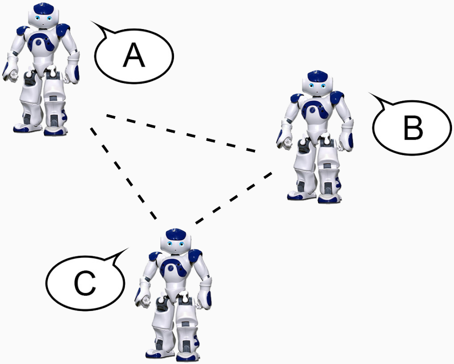
Introduction

Multi-Agent Systems



Sources of uncertainty:

- states
- actions
- behaviour



Model-free methods:

- E.g. regret, policy gradient, model-free RL
- *Does not address behaviour uncertainty*

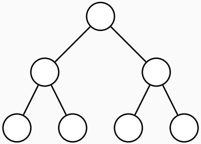
Agent Modelling

Model-free methods:

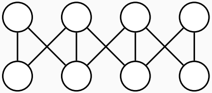
- E.g. regret, policy gradient, model-free RL
- *Does not address behaviour uncertainty*

Model-based methods:

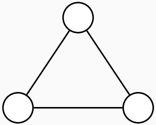
- Learn model of agent behaviour during interaction, e.g.



Decision tree



Neural network



State machine

- Use model to plan own actions

Agent Modelling

Why agent modelling?

- Generalise observations to unseen situations
- Plan into the future (e.g. guided exploration, risk control)
- But...

Agent Modelling

Why agent modelling?

- Generalise observations to unseen situations
- Plan into the future (e.g. guided exploration, risk control)
- But...

Problem: no model criticism

Does not check validity of model during interaction

May use *incorrect model* without ever realising it

Agent Modelling – Example

Simple example:

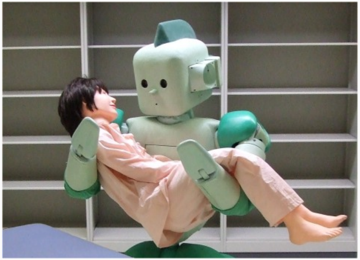
- Rock-Paper-Scissors
- Human plays R-P-S-R-P-S-...

Model human as fixed distribution:

- Limit model is $< \frac{1}{3}, \frac{1}{3}, \frac{1}{3} >$
- Expected payoff with correct model is 1
- Expected payoff with learned model is 0

Robot never realises that model is wrong!

Agent Modelling – Example



Complex examples:

- elderly support
- user interfaces
- electronic markets

What can go wrong?

- In general, anything
- Wrong models can make wrong predictions
- Wrong predictions can lead to bad actions

Belief and Truth

Model is effective hypothesis (*belief*) of agent

- Hypothesis can be *false*
- But: model not treated as hypothesis

Idea: learn beliefs over multiple models

$$\underbrace{\boxed{\theta_1 \mid \theta_1 \mid \ldots \mid \theta_n}}_{\Pr(\theta_x|H^t)}$$

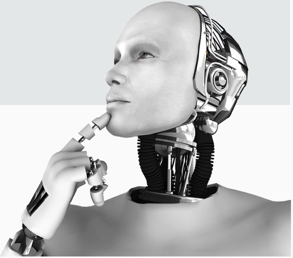
Same problem:

- $\Pr(\theta_x|H^t)$ is relative likelihood, not absolute truth
- Models may still be wrong

Belief and Truth

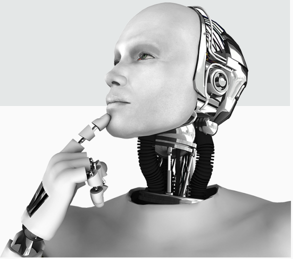
We need agent to do both:

- Construct hypothesis of behaviour
- Contemplate truth of hypothesis



We need agent to do both:

- Construct hypothesis of behaviour
- Contemplate **truth** of hypothesis



Allows agent to...

- Reject model
- Change assumptions
- Change modelling method
- Get *better* model

  – or –

- Resort to *safe* policy with no/minimal model

Behavioural Hypothesis Testing

Behavioural Hypothesis Testing

Model is hypothesis because:

- true or false
- testable

Natural question:

Given hypothesis π_j^* for agent j and history H^t,
does j really behave according to π_j^*?

Behavioural Hypothesis Testing – Example

t	(a_1^t, a_2^t)	π_2^*
1	(R, P)	$\langle .3, .1, .6 \rangle$
2	(S, R)	$\langle .2, .3, .5 \rangle$
3	(P, S)	$\langle .7, .1, .2 \rangle$
4	(P, S)	$\langle .0, .4, .6 \rangle$
5	(R, P)	$\langle .4, .2, .4 \rangle$

Behavioural Hypothesis Testing

Natural to compute some score from table:

- e.g. empirical frequency (Conitzer and Sandholm, 2007; Foster and Young, 2003)
- But: when is scoring scheme sufficient?
- But: how to choose threshold parameter for score?

Behavioural Hypothesis Testing

Natural to compute some score from table:

- e.g. empirical frequency
  (Conitzer and Sandholm, 2007; Foster and Young, 2003)
- But: when is scoring scheme sufficient?
- But: how to choose threshold parameter for score?

Proposed solution: Frequentist hypothesis test (p-value)

- Allow for multiple scoring criteria in test statistic
- Significance level α invariant of scoring scheme

Each agent i has behaviour $\pi_i \in \Pi_i$

- $\pi_i(H^t) \in \Delta(A_i)$
- A_i is action space for agent i
- $H^t = (s^0, a^0, s^1, a^1, ..., s^t)$ is history
- s^τ is signal/state observed at time τ
- $a^\tau = (a_1^\tau, ..., a_m^\tau)$ is tuple of actions taken at time τ

Two-Sample Problem

We control i and observe j

- π_j is true behaviour of j
- π_j^* is hypothesised behaviour of j
- Question: $\pi_j^* = \pi_j$?

Two-Sample Problem

We control i and observe j

- π_j is true behaviour of j
- π_j^* is hypothesised behaviour of j
- Question: $\pi_j^* = \pi_j$?

Cannot answer directly since π_j unknown, but

- We know $\mathbf{a}_j^t = (a_j^0, ..., a_j^{t-1})$ from H^t
- Can generate $\hat{\mathbf{a}}_j^t = (\hat{a}_j^0, ..., \hat{a}_j^{t-1})$ using π_j^*
- Two-sample problem: were $\mathbf{a}_j^t$ and $\hat{\mathbf{a}}_j^t$ generated by π_j^*?

Frequentist Hypothesis Test

Compute p-value:

$$p = P\left(\ |T(\tilde{a}_j^t, \hat{a}_j^t)|\ \geq\ |T(a_j^t, \hat{a}_j^t)|\ \right)$$

$$\tilde{a}_j^t \sim \left(\pi_j^*(H^0), \ldots, \pi_j^*(H^{t-1})\right)$$

Null-assumption: $\pi_j^* = \pi_j$

Reject π_j^* if p below some *significance level* $\alpha \in [0, 1]$

Test Statistic

Test statistic:

$$
\begin{aligned}
T(\tilde{\mathbf{a}}_j^t, \hat{\mathbf{a}}_j^t) &= \frac{1}{t}\sum_{\tau=1}^{t} T_\tau(\tilde{\mathbf{a}}_j^\tau, \hat{\mathbf{a}}_j^\tau) \\
T_\tau(\tilde{\mathbf{a}}_j^\tau, \hat{\mathbf{a}}_j^\tau) &= \sum_{k=1}^{K} w_k \left(z_k(\tilde{\mathbf{a}}_j^\tau, \pi_j^*) - z_k(\hat{\mathbf{a}}_j^\tau, \pi_j^*) \right)
\end{aligned}
$$

$w_k \in \mathbb{R}$ is weight for *score function* $z_k \in Z$

Intuition: $z_k(\tilde{\mathbf{a}}_j^\tau, \pi_j^*)$ *likelihood* that π_j^* produced $\tilde{\mathbf{a}}_j^\tau$

Example Score Functions

$$z_1(\mathsf{a}_j^t, \pi_j^*) = \frac{1}{t}\sum_{\tau=0}^{t-1} \frac{\pi_j^*(H^\tau)[a_j^\tau]}{\max_{a_j} \pi_j^*(H^\tau)[a_j]}$$

Example Score Functions

$$z_1(\mathsf{a}_j^t, \pi_j^*) \quad = \quad \frac{1}{t}\sum_{\tau=0}^{t-1} \frac{\pi_j^*(H^\tau)[a_j^\tau]}{\max_{a_j} \pi_j^*(H^\tau)[a_j]}$$

$$z_2(\mathsf{a}_j^t, \pi_j^*) \quad = \quad \frac{1}{t}\sum_{\tau=0}^{t-1} 1 - \mathbb{E}_{a_j \sim \pi_j^*(H^\tau)} \left| \pi_j^*(H^\tau)[a_j^\tau] - \pi_j^*(H^\tau)[a_j] \right|$$

Example Score Functions

$$z_1(\mathsf{a}_j^t, \pi_j^*) = \frac{1}{t}\sum_{\tau=0}^{t-1} \frac{\pi_j^*(H^\tau)[a_j^\tau]}{\max_{a_j} \pi_j^*(H^\tau)[a_j]}$$

$$z_2(\mathsf{a}_j^t, \pi_j^*) = \frac{1}{t}\sum_{\tau=0}^{t-1} 1 - \mathbb{E}_{a_j \sim \pi_j^*(H^\tau)} \left| \pi_j^*(H^\tau)[a_j^\tau] - \pi_j^*(H^\tau)[a_j] \right|$$

$$z_3(\mathsf{a}_j^t, \pi_j^*) = \sum_{a_j \in A_j} \min \left[\frac{1}{t}\sum_{\tau=0}^{t-1}[a_j^\tau = a_j]_1 \,,\, \frac{1}{t}\sum_{\tau=0}^{t-1} \pi_j^*(H^\tau)[a_j] \right]$$

Learning the Test Distribution

Can show that test statistic eventually normal, but:

- shaped gradually over time
- initially skewed

Need special distribution to capture dynamics:

- Skew-normal distribution (Azzalini, 1985)

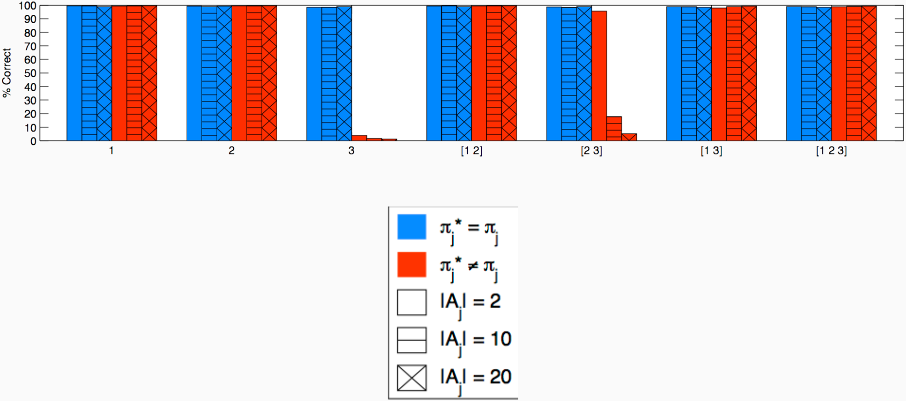
$$f(x \mid \xi, \omega, \beta) = \frac{2}{\omega}\,\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\beta\left(\frac{x-\xi}{\omega}\right)\right)$$

- ϕ and Φ are standard normal PDF and CDF
- Learn parameters ξ, ω, β during interaction

Experiments: Random Behaviours

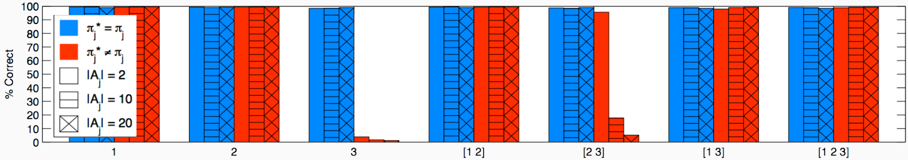
π_i, π_j, π_j^*: random action distribution in each time step

Tested all combinations of score functions z_1, z_2, z_3

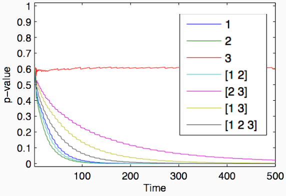


Experiments: Random Behaviours

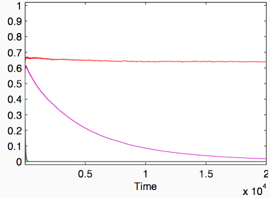
π_i, π_j, π_j^*: random action distribution in each time step



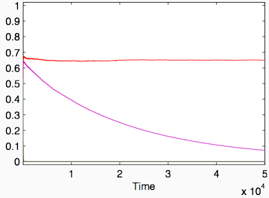
Score combination can "heal" convergence:



$|A_j| = 2$



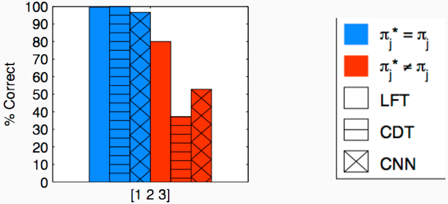
$|A_j| = 10$



$|A_j| = 20$

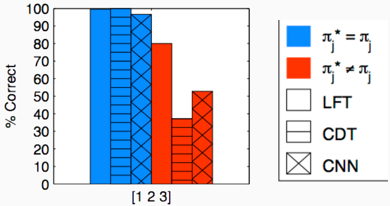
Experiments: Adaptive Behaviours

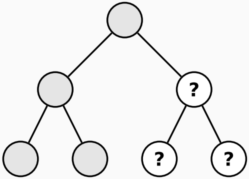
π_i, π_j, π_j^*: behaviour from same adaptive class (LFT, CDT, CNN)



Experiments: Adaptive Behaviours

π_i, π_j, π_j^*: behaviour from same adaptive class (LFT, CDT, CNN)





Limitation:

Does not probe specific aspects of hypothesis!

Publications

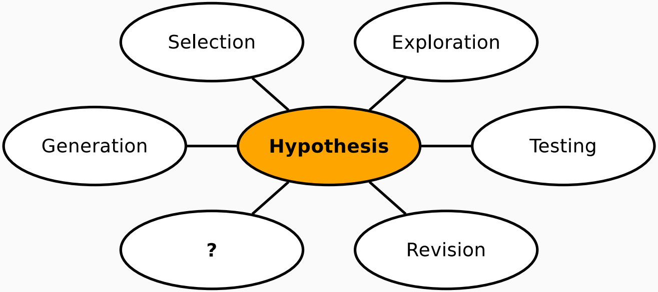
- S. Albrecht, J. Crandall, and S. Ramamoorthy.
  Belief and truth in hypothesised behaviours.
  Artificial Intelligence, 235:63–94, 2016.
- S. Albrecht and S. Ramamoorthy.
  Are you doing what I think you are doing? Criticising uncertain agent models.
  In *31st Conference on Uncertainty in Artificial Intelligence*, pages 52–61, 2015.

The Future

The Future

Testing is only part of bigger picture…

- Need hypothesis *“contemplation”*



The Future

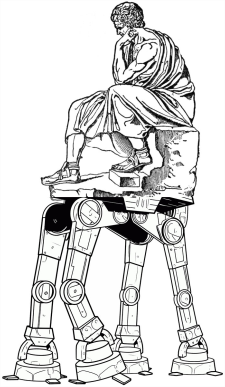
Testing is only part of bigger picture...

- Need hypothesis *"contemplation"*

The Future

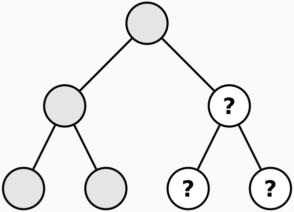
Testing is only part of bigger picture...

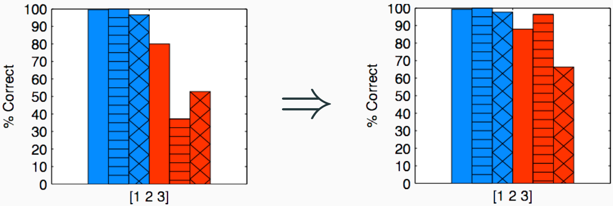
- Need hypothesis *“contemplation”*



Exploration:

- How and when to explore aspects of hypothesis?





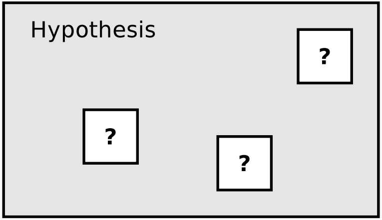
using random exploration

Revision:

- How to revise and improve aspects of hypothesis?

Example:

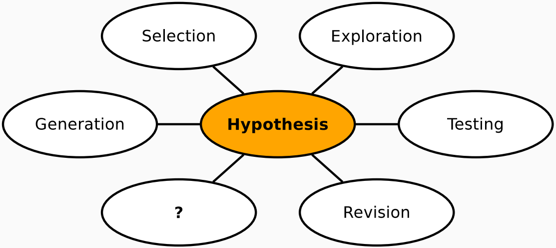
- Hypothesis is reinforcement learner
- How to revise
  … *learning rate?*
  … *exploration rate?*
  … *discount rate?*



The Future

Individual pieces of puzzle exist

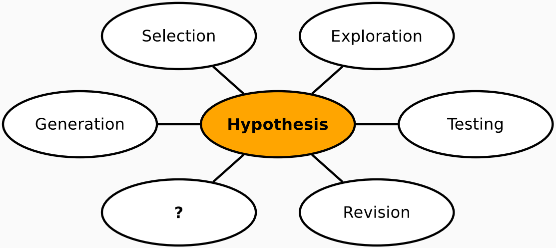
- Need integration into complete solution
- Important, feasible, and timely
- Relevant in all areas of AI



The Future

Challenges:

- Complexity, soundness, completeness, etc.
- Contemplate usefulness, not just correctness
- Can agent learn on its own how to contemplate?



Thank you