Learning to Distinguish Between Belief and Truth

Stefano V. Albrecht
June 15, 2016

Department of Computer Science
The University of Texas at Austin
Introduction

Behavioural Hypothesis Testing

The Future
Introduction
Multi-Agent Systems
Sources of uncertainty:

- states
- actions
- behaviour
Agent Modelling

Model-free methods:

- E.g. regret, policy gradient, model-free RL
- *Does not address behaviour uncertainty*
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...
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 \textit{false}
- But: model not treated as hypothesis

\textbf{Idea:} learn beliefs over multiple models

\[
\begin{array}{c|c|c|c}
\theta_1 & \theta_1 & \ldots & \theta_n \\
\hline
\end{array}
\]

\[Pr(\theta_x|H^t)\]

\textbf{Same problem:}
- \(Pr(\theta_x|H^t)\) is relative likelihood, not absolute truth
- Models may still be \textit{wrong}
Belief and Truth

We need agent to do both:

- Construct hypothesis of behaviour
- Contemplate **truth** of hypothesis
Belief and Truth

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 $\pi_j^*$ for agent $j$ and history $H^t$, does $j$ really behave according to $\pi_j^*$?
## Behavioural Hypothesis Testing – Example

<table>
<thead>
<tr>
<th>$t$</th>
<th>$(a^t_1, a^t_2)$</th>
<th>$\pi^*_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$(R, P)$</td>
<td>$\langle .3, .1, .6 \rangle$</td>
</tr>
<tr>
<td>2</td>
<td>$(S, R)$</td>
<td>$\langle .2, .3, .5 \rangle$</td>
</tr>
<tr>
<td>3</td>
<td>$(P, S)$</td>
<td>$\langle .7, .1, .2 \rangle$</td>
</tr>
<tr>
<td>4</td>
<td>$(P, S)$</td>
<td>$\langle .0, .4, .6 \rangle$</td>
</tr>
<tr>
<td>5</td>
<td>$(R, P)$</td>
<td>$\langle .4, .2, .4 \rangle$</td>
</tr>
</tbody>
</table>
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?
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 $\alpha$ 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
- $s^\tau$ is signal/state observed at time $\tau$
- $a^\tau = (a^\tau_1, ..., a^\tau_m)$ is tuple of actions taken at time $\tau$
Two-Sample Problem

We control $i$ and observe $j$

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

We control $i$ and observe $j$

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

Cannot answer directly since $\pi_j$ unknown, but

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

Compute \( p \)-value:

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

\[\hat{a}_j^t \sim \left( \pi_j^*(H^0), ..., \pi_j^*(H^{t-1}) \right)\]

Null-assumption: \( \pi_j^* = \pi_j \)

Reject \( \pi_j^* \) if \( p \) below some **significance level** \( \alpha \in [0, 1] \)
Test Statistic

Test statistic:

\[
T(\tilde{a}_j^T, \hat{a}_j^T) = \frac{1}{t} \sum_{\tau=1}^{t} T_{\tau}(\tilde{a}_j^T, \hat{a}_j^T)
\]

\[
T_{\tau}(\tilde{a}_j^T, \hat{a}_j^T) = \sum_{k=1}^{K} w_k \left( z_k(\tilde{a}_j^T, \pi_j^*) - z_k(\hat{a}_j^T, \pi_j^*) \right)
\]

\(w_k \in \mathbb{R}\) is weight for score function \(z_k \in Z\)

Intuition: \(z_k(\tilde{a}_j^T, \pi_j^*)\) likelihood that \(\pi_j^*\) produced \(\tilde{a}_j^T\)
Example Score Functions

\[ z_1(a_j^t, \pi_j^*) = \frac{1}{t} \sum_{\tau=0}^{t-1} \frac{\pi_j^*(H^\tau)[a_j^\tau]}{\max_{a_j} \pi_j^*(H^\tau)[a_j]} \]
Example Score Functions

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

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

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

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

\[ z_3(a_j^t, \pi_j^*) = \min_{a_j \in A_j} \left[ \frac{1}{t} \sum_{\tau=0}^{t-1} [a_j^\tau = a_j], \frac{1}{t} \sum_{\tau=0}^{t-1} \pi_j^*(H^\tau)[a_j] \right] \]
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)

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

- \( \phi \) and \( \Phi \) are standard normal PDF and CDF
- Learn parameters \( \xi, \omega, \beta \) during interaction
Experiments: Random Behaviours

$\pi_i$, $\pi_j$, $\pi_j^*$: random action distribution in each time step

Tested all combinations of score functions $z_1, z_2, z_3$
Experiments: Random Behaviours

$\pi_i, \pi_j, \pi_j^*$: random action distribution in each time step

Score combination can “heal” convergence:

$|A_j| = 2$  
$|A_j| = 10$  
$|A_j| = 20$
Performing tests on adaptive behaviors \( \pi_i, \pi_j, \pi_j^* \): behavior from same adaptive class (LFT, CDT, CNN)
Experiments: Adaptive Behaviours

$\pi_i, \pi_j, \pi_j^*$: behaviour from same adaptive class (LFT, CDT, CNN)

**Limitation:**

Does not probe specific aspects of hypothesis!
**Belief and truth in hypothesised behaviours.**

S. Albrecht and S. Ramamoorthy. 
**Are you doing what I think you are doing? Criticising uncertain agent models.**
The Future
Testing is only part of bigger picture...

- Need hypothesis "contemplation"
The Future

Testing is only part of bigger picture...

• Need hypothesis “contemplation”
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*?
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