Computational models of human inductive learning

Josh Tenenbaum, Charles Kemp, Tom Griffiths
What is this about?

Starting idea: it's interesting to study mind's ability to build rich models of the world from sparse data.

So,

*ideas from cognitive science +

*ideas from developmental psychology +

*statistical methods =

a proposal for a computational model of human inductive learning
Why?

...and because they are interested in how humans learn, rather...
Seriously, why?

Because it would be great to understand how a child is able to understand the concept of a “horse” from just a few examples for instance
And?

Because it might bring new insights in machine learning.

Because such a computational model could be a fertile base for new applications.
Everyday inductive leaps

How can people learn so much about the world from such limited evidence?

Kinds of objects and their properties
The meanings of words, phrases, and sentences
Cause-effect relations
The beliefs, goals and plans of other people
Social structures, conventions, and rules
The solution

Strong prior knowledge (inductive bias).
The solution

Strong prior knowledge (inductive bias).
How does background knowledge guide learning from sparsely observed data?
What form does the knowledge take, across different domains and tasks?
How is that knowledge itself acquired?
How can inductive biases be so strong yet so flexible?

The goal is a computational framework for answering these questions.
The approach: from *statistics* to *intelligence*

1. How does background knowledge guide learning from sparsely observed data?
   Bayesian inference, with priors based on background knowledge.

2. What form does background knowledge take, across different domains and tasks?
   Probabilities defined over structured representations: graphs, grammars, predicate logic, schemas, theories.

3. How is background knowledge itself acquired?
   Hierarchical Bayesian models, with inference at multiple levels of abstraction.

4. How can inductive biases be so strong yet so flexible?
   Nonparametric models, growing in complexity as the data require.
Outline

Bags of marbles
Word learning
Hierarchical Bayes model
Causal learning
Property induction
Outline

Bags of marbles
Word learning
Hierarchical Bayes model
Causal learning
Property induction
Learning about feature variability

Marbles of different colors:  🟠🟢🔴🔵🟢 ...

✔️  ✔️  ✔️  ✔️  ✔️  🟠?

✔️  ✔️  ✔️  ✔️  ✔️  🟠?
Learning about feature variability

Marbles of different colors: 

...
A hierarchical Bayesian model

Level 1: Bag proportions
- mostly red
- mostly yellow
- mostly brown
- mostly blue
- ... (other colors)

Level 2: Bags in general
- mostly yellow
- mostly green

Color varies across bags but not much within bags.
A hierarchical Bayesian model

Prior expectations on bags in general

Level 2: Bags in general

Level 1: Bag proportions

Data

\[ \lambda \]

\[ \alpha \sim \text{Exponential}(\lambda) \]

\[ \beta \sim \text{Dirichlet}(1) \]

\[ \theta^i \sim \text{Dirichlet}(\alpha \beta) \]

\[ y^i \sim \text{Multinomial}(\theta^i) \]
A hierarchical Bayesian model

Level 1: Bag proportions

Level 2: Bags in general

Level 3: Prior expectations on bags in general

Data

Prior expectations on bags in general

Bag 1 is mostly red

\[ \alpha \sim \text{Exponential}(\lambda) \]

\[ \beta \sim \text{Dirichlet}(1) \]

\[ \theta^i \sim \text{Dirichlet}(\alpha \beta) \]

\[ y^i \sim \text{Multinomial}(\theta^i) \]
A hierarchical Bayesian model

Level 1: Bag proportions

Level 2: Bags in general

Prior expectations on bags in general

"Color varies across bags but not much within bags"

Level 1: Bag proportions

\[ \theta^1, \theta^2, \theta^3, \theta^4, \ldots, \theta^n \]

Level 2: Bags in general

\[ \alpha, \beta \]

\[ \alpha \sim \text{Exponential}(\lambda) \]

\[ \beta \sim \text{Dirichlet}(1) \]

\[ \theta^i \sim \text{Dirichlet}(\alpha\beta) \]

\[ y^i \sim \text{Multinomial}(\theta^i) \]
Outline

Bags of marbles

Word learning

Hierarchical Bayes model

Causal learning

Property induction
The “shape bias” in word learning
(Landau, Smith, Jones 1988)

This is a dax. Show me the dax…

English-speaking children show the shape bias at 24 months, but not at 20 months.
The shape bias is a useful inductive constraint: majority of early words are labels for object categories, and shape may be the best cue to object category membership.
Learning the shape bias

Assume independent Dirichlet-multinomial models for each dimension. Should learn:
Shape varies across categories but not within categories.
Texture, color, size vary within categories.
Learning the shape bias

This is a dax.

<table>
<thead>
<tr>
<th>Category</th>
<th>1 1 2 2 3 3 4 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape</td>
<td>1 1 2 2 3 3 4 4</td>
</tr>
<tr>
<td>Texture</td>
<td>1 2 3 4 5 6 7 8</td>
</tr>
<tr>
<td>Color</td>
<td>1 2 3 4 5 6 7 8</td>
</tr>
<tr>
<td>Size</td>
<td>1 2 1 2 1 2 1 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
</tr>
<tr>
<td>?</td>
</tr>
<tr>
<td>?</td>
</tr>
<tr>
<td>?</td>
</tr>
</tbody>
</table>

| 5 |
| 5 |
| 6 |
| 6 |

| 9 |
| 10 |
| 9 |
| 10 |
| 9 |

| 1 |
| 1 |
| 1 |
| 1 |

Show me the dax.
Learning to transfer selectively

Let $z_i$ be the ontological kind of category $i$.

Given $z$, we could learn a separate Dirichlet-multinomial model for each ontological kind:

Variability in solidity, shape, material within kind 1

Variability in solidity, shape, material within kind 2

“dax” shape

“toof” material
Learning to transfer selectively

Chicken-and-egg problem: We don’t know the partition \( z \) into ontological kinds.

The input:

Solution: Define a nonparametric prior over this partition.

\[
\begin{align*}
\left| n_{\text{cat}} \right| & \sim \text{CRP}(\gamma) \\
\alpha^k & \sim \text{Exponential}(\lambda) \\
\beta^k & \sim \text{Dirichlet}(1) \\
\theta^i & \sim \text{Dirichlet}(\alpha^{z_i} \beta^{z_i}) \\
y^i & \sim \text{Multinomial}(\theta^i)
\end{align*}
\]
Outline

Bags of marbles
Word learning
Hierarchical Bayes model
Causal learning
Property induction
Hierarchical Bayesian Framework

\[ P(S, F | D) \propto P(D | S) P(S | F) P(F) \]
HBF - Generic case

3. Inferring form

\[ P(T_1 | D, T_2) = \frac{P(D|T_1)P(T_1|T_2)}{P(D|T_2)}. \]

\[ P(D|T_1) = \prod_{i=1}^{L} P(d_i|T_1). \]

2. Inferring structure

\[ P(d|T_0) = \prod_{i=1}^{M} P(x^{(i)}_{\text{obs}}|T_0). \]

\[ P(T_0|d, T_1) = \frac{P(d|T_0)P(T_0|T_1)}{P(d|T_1)}, \]

\[ P(d|T_1) = \sum_{T_0 \in H_1} P(d|T_0)P(T_0|T_1). \]

1. Inferring unobserved data

\[ P(x_{\text{unobs}}|x_{\text{obs}}, T_0) = \frac{P(x_{\text{obs}}|x_{\text{unobs}}, T_0)P(x_{\text{unobs}}|T_0)}{P(x_{\text{obs}}|T_0)} \]

\[ P(x_{\text{obs}}|T_0) = \sum_{x_{\text{unobs}} \in H_0} P(x_{\text{obs}}|x_{\text{unobs}}, T_0)P(x_{\text{unobs}}|T_0). \]
HBF – (some) issues

Intractable to compute all probabilities for more than 2-3 levels of the hierarchy or for large systems.

=> Solution: stochastic sampling or search for the most probable hypothesis

How could the direct learning of abstract structures (from an expert) be modeled?

=> Open question
Outline

Bags of marbles
Word learning
Hierarchical Bayes model
Causal learning
Property induction
Bayesian network

Learning causal relations

Data

Patient 1: Stressful lifestyle
  Chest Pain
Patient 2: Smoking
  Coughing
Patient 3: Working in factory
  Chest Pain
  ...
Learning causal relations

Abstract theory

Three types of variables: Behavior(X), Disease(X), Symptom(X)
Behaviors can cause Diseases
Diseases can cause Symptoms

Principles

Classes: \{R, D, S\} (Risks, Diseases, Symptoms)
Causal laws: \( R \rightarrow D, \ D \rightarrow S \)

Bayesian network

Structure

Data

Patient 1: Stressful lifestyle
Chest Pain
Patient 2: Smoking
Coughing
Patient 3: Working in factory
Chest Pain
...
Figure 3. An analogy between multiple levels of structure in (a) knowledge of language and (b) causal knowledge. Each level generates structures at the level below, thereby establishing necessary constraints on the hypothesis space for inductive inference.
Representation - graph grammars

Node classes:

<table>
<thead>
<tr>
<th>Class</th>
<th>Symbol</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior</td>
<td>B</td>
<td>open</td>
</tr>
<tr>
<td>Disease</td>
<td>D</td>
<td>open</td>
</tr>
<tr>
<td>Symptom</td>
<td>S</td>
<td>open</td>
</tr>
</tbody>
</table>

Class graph:

\[ B \rightarrow D \rightarrow S \]

Generative model:

1. Generate nodes in each class.

\[ N_B \sim \text{PowerLaw}(\alpha_B) \]
\[ N_D \sim \text{PowerLaw}(\alpha_D) \]
\[ N_S \sim \text{PowerLaw}(\alpha_S) \]

2. Generate causal relations between pairs of nodes.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Relation</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b \in B, d \in D )</td>
<td>( b \rightarrow d )</td>
<td>( \beta_{BD} )</td>
</tr>
<tr>
<td>( d \in D, s \in S )</td>
<td>( d \rightarrow s )</td>
<td>( \beta_{DS} )</td>
</tr>
</tbody>
</table>

*Figure 2.* A graph schema \( G_{\text{Dis}} \) for networks of diseases, their causes and their effects.
Graph grammar for magnets world

(a) Node classes:

<table>
<thead>
<tr>
<th>Class</th>
<th>Symbol</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position of a magnet</td>
<td>$M$</td>
<td>open</td>
</tr>
<tr>
<td>Position of a magnetic object</td>
<td>$T$</td>
<td>open</td>
</tr>
<tr>
<td>Position of a non-magnetic object</td>
<td>$U$</td>
<td>open</td>
</tr>
</tbody>
</table>

(b) Class graph:

Figure 5. (a) A graph schema $G_{Mag}$ describing the effects of magnets on other objects. (b) Causal networks sampled from the grammar.
Difficulty of learning graph grammars

Figure 9. (a) Class graphs and sample networks representing the four graph schemas explored in the experiments of Tenenbaum and Niyogi (2003). (b) The evidence for a theory based on a graph schema increases as learners encounter more objects exhibiting causal relations consistent with that schema, but at a different rate for different graph schemas. Human learners demonstrate the same ordering in the difficulty of learning these graph schemas.
Outline

Bags of marbles
Word learning
Hierarchical Bayes model
Causal learning
Property induction
Hierarchical Bayesian Framework again

$F$: form

$S$: structure

$D$: data
Discovering structural forms

“Great chain of being”

Linnaeus

Ostrich  Robin  Crocodile  Snake  Turtle  Bat  Orangutan  Angel  God

Ostrich  Robin  Crocodile  Snake  Turtle  Bat  Orangutan
Property induction example

'Animals' set
106 features
33 species
Experiments with tree, chain, ring and partition forms.
40 minutes simulation in Matlab (code from Tenenbaum's website) – see following slides
A tree form for 'Animals' set
A chain form for 'Animals' set
A ring form for 'Animals' set
A partition form for 'Animals' set
Central themes

Bayesian inference.
Bayesian models rely on a prior, and can therefore incorporate structured background knowledge.

Structured representations.
Structured representations provide an inductive bias that is crucial when learning from sparse or noisy data.

Multiple levels of abstraction.
Hierarchical models explain how structured representations can be acquired.
Resources

Majority of slides from the original presentations and all figures from the articles / presentations.

Bayesian models of human learning - video lecture ICML 2007


“Intuitive Theories as Grammars for Causal Inference” and “Two Proposals for Causal Grammars” - Tenenbaum, Griffiths 2007