CHUNK LEARNING:
HIERARCHICAL PAIR-WISE ASSOCIATIVE?
CHUNK LEARNING: HIERARCHICAL PAIR-WISE ASSOCIATIVE?

Gergő Orbán
CHUNK LEARNING:
HIERARCHICAL PAIR-WISE ASSOCIATIVE?

Gergő Orbán
CHUNK LEARNING: HIERARCHICAL PAIR-WISE ASSOCIATIVE?

Gergő Orbán
CHUNK LEARNING:
HIERARCHICAL PAIR-WISE ASSOCIATIVE?
CHUNK LEARNING:
HIERARCHICAL PAIR-WISE ASSOCIATIVE?

Gergő Orbán
CHUNK LEARNING:
HIERARCHICAL PAIR-WISE ASSOCIATIVE?
CHUNK LEARNING: HIERARCHICAL PAIR-WISE ASSOCIATIVE?
CHUNK LEARNING: HIERARCHICAL PAIR-WISE ASSOCIATIVE?

Probabilistic models of learning and memory — Structural learning

CEU, Budapest, 22-26 June 2009

http://www.eng.cam.ac.uk/~m.lengyel
familiarization
‘pay attention’
familiarization
‘pay attention’
VISUAL PATTERN LEARNING

familiarization
‘pay attention’

test
‘which one looks more familiar?’
VISUAL PATTERN LEARNING

familiarization
‘pay attention’

test
‘which one looks more familiar?’
VISUAL PATTERN LEARNING

familiarization
‘pay attention’

inventory

‘which one looks more familiar?’

test

[Images of visual patterns]
VISUAL PATTERN LEARNING

familiarization
‘pay attention’

inventory

‘which one looks more familiar?’

test
VISUAL PATTERN LEARNING

familiarization
‘pay attention’

inventory

‘which one looks more familiar?’

test

Orbán & al, 2008
VISUAL PATTERN LEARNING

familiarization
‘pay attention’

\[ D = x_1, x_2, \ldots, x_n \quad \rightarrow \quad \hat{P}(x) = \int \hat{P}(x|M) \hat{P}(M|D) \, dM \quad \rightarrow \quad \hat{P}(x_A) \text{ vs. } \hat{P}(x_B) \]

inventory

‘which one looks more familiar?’

\[ \hat{P}(M|D) \]

test

\[ \hat{P}(x) \]

Orbán & al, 2008

Probabilistic models of learning and memory — Structural learning

CEU, Budapest, 22-26 June 2009

http://www.eng.cam.ac.uk/~m.lengyel
VISUAL PATTERN LEARNING

familiarization
‘pay attention’

inventory

test
‘which one looks more familiar?’

\[ D = x_1, x_2, \ldots, x_n \rightarrow \hat{P}(x) = \int \hat{P}(x|M) \hat{P}(M|D) dM \rightarrow \hat{P}(x_A) \text{ vs. } \hat{P}(x_B) \]

how do humans learn a statistical model of their environment?
VISUAL PATTERN LEARNING

familiarization
‘pay attention’

inventory

test
‘which one looks more familiar?’

\[ D = x_1, x_2, \ldots, x_n \rightarrow \hat{P}(x) = \int \hat{P}(x|M) \hat{P}(M|D) dM \rightarrow \hat{P}(x_A) \text{ vs. } \hat{P}(x_B) \]

how do humans learn a statistical model of their environment?

• associative learning (fitting 2nd order max-entropy model)
VISUAL PATTERN LEARNING

how do humans learn a statistical model of their environment?

- associative learning (fitting 2\textsuperscript{nd} order max-entropy model)
- Bayesian model selection (inferring hidden causal structure)
ALTERNATIVE THEORIES
ALTERNATIVE THEORIES

associative learning
ALTERNATIVE THEORIES

associative learning
ALTERNATIVE THEORIES

associative learning

Bayesian learning
ALTERNATIVE THEORIES

associative learning

Bayesian learning
ALTERNATIVE THEORIES

associative learning

Boltzmann machine
+ Gaussian Markov random field

Bayesian learning

sigmoid belief network
+ product of (conditional) Gaussian experts
MULTIPLE EXPERIMENTS

baseline

inventory

test type

basic

vs.
MULTIPLE EXPERIMENTS

baseline

inventory

test type

basic

% correct

humans
associative learner
Bayesian learner

basic

vs.
MULTIPLE EXPERIMENTS

baseline vs. frequency-balanced

inventory

rare vs. frequent

basic vs. frequency-balanced

test type

% correct

humans
associative learner
Bayesian learner

basic vs. frequency-balanced

test performance

% correct
MULTIPLE EXPERIMENTS

- **inventory**
  - baseline
  - frequency-balanced
    - rare
    - frequent

- **test type**
  - basic vs. frequency-balanced
  - triplet
    - basic vs. embedded

- **test performance**
  - % correct
    - basic
    - frequency-balanced
    - basic vs. embedded

Humans, associative learner, Bayesian learner
MULTIPLE EXPERIMENTS

inventory

baseline

frequency-balanced

rare

frequent

test type

basic vs. frequency-balanced

vs.

basic vs. embedded

basic vs. embedded

test performance

% correct

humans

associative learner

Bayesian learner

basic

frequency-balanced

triplet

quadruple

http://www.eng.cam.ac.uk/~m.lengyel
MULTIPLE EXPERIMENTS

inventory

baseline

frequent

rare

frequency-balanced

triplet

quadruple

test type

basic vs. frequency-balanced

embedding

human associative learner Bayesian learner

humans

associative learner

Bayesian learner

test performance

% correct

basic

frequency-balanced

basic embedded

basic quad pair

embedded pair triplet

humans

associative learner

Bayesian learner

humans

associative learner

Bayesian learner

humans

associative learner

Bayesian learner

humans

associative learner

Bayesian learner

humans

associative learner

Bayesian learner

humans

associative learner

Bayesian learner

humans

associative learner

Bayesian learner

Orbán & al, 2008

Probabilistic models of learning and memory — Structural learning

CEU, Budapest, 22-26 June 2009

http://www.eng.cam.ac.uk/~m.lengyel
EMBEDDED RESULTS

Probabilistic models of learning and memory — Structural learning

http://www.eng.cam.ac.uk/~m.lengyel

CEU, Budapest, 22-26 June 2009  http://www.eng.cam.ac.uk/~m.lengyel
chunk-based explanation

\[ P_{\text{chunk}} = p \cdot (1 - \delta)^n \cdot \delta^{(N-n)} \]
EMBEDDED RESULTS

chunk-based explanation

\[ P_{\text{chunk}} = p \cdot (1 - \delta)^n \cdot \delta^{(N-n)} \]

noise-based explanation

\[ P_{\text{noise}} = (1 - p) \cdot \epsilon^n \]

\[ P = \begin{cases} 1 - \delta & \text{if } y = \text{chunk} \\ 1 - \delta & \text{if } y = \text{shapes} \end{cases} \]

\[ \epsilon \]

\[ x_1 \quad \ldots \quad x_n \quad \ldots \quad x_N \]

\[ \epsilon \quad \epsilon \quad \epsilon \]
EMBEDDED RESULTS

chunk-based explanation

\[ P_{\text{chunk}} = p \cdot (1 - \delta)^n \cdot \delta^{(N-n)} \]

noise-based explanation

\[ P_{\text{noise}} = (1 - p) \cdot \epsilon^n \]

\[ p \]

\[ y \]

\[ 1 - \delta \]

\[ 1 - \delta \]

\[ 1 - \delta \]

\[ x_1 \]

\[ \ldots \]

\[ x_n \]

\[ \ldots \]

\[ x_N \]

\[ \epsilon \]

\[ \epsilon \]

\[ \epsilon \]

\[ \epsilon \]

\[ \epsilon \]

Orbán & al, 2008

0

0.05

0.1

0.15

0.2

0.25

1

2

3

4

5

6

\[ N = 2 \]

\[ N = 3 \]

\[ N = 4 \]

\[ N = 5 \]

\[ N = 6 \]
ASSOCIATIVE VS. BAYESIAN LEARNING

inventory:
ASSOCIATIVE VS. BAYESIAN LEARNING

inventory:
ASSOCIATIVE VS. BAYESIAN LEARNING

inventory:
ASSOCIATIVE VS. BAYESIAN LEARNING

inventory:

1st order statistic: shape frequencies

1:6

1:2

1:6

2:6

3 × 1/6

1/6 + 2/6
ASSOCIATIVE VS. BAYESIAN LEARNING

inventory:

1st order statistic: shape frequencies

- 3 × 1/6

2nd order statistic: pairwise correlations

- both present: 2 × 1/6
- both absent: 2/6
- one present: 2 × 1/6

- 1/6 + 2/6

- 2/6
- 2 × 1/6
- 2 × 1/6
ASSOCIATIVE VS. BAYESIAN LEARNING

inventory:

1\textsuperscript{st} order statistic: shape frequencies

2\textsuperscript{nd} order statistic: pairwise correlations

both present
both absent
one present

3 × 1/6
2 × 1/6
2 × 1/6

1/6 + 2/6

2/6
2 × 1/6
2 × 1/6

test performance:

% correct

humans
associative learner
Bayesian learner

Orbán & al, 2008
QUANTITATIVE COMPARISON

Bayesian learner

\[ r = 0.88 \ (p < 0.0002) \]

Orbán & al, 2008

associative learner

\[ r = 0.71 \ (p < 0.01) \]
QUANTITATIVE COMPARISON

Bayesian learner

\[ r = 0.88 \ (p < 0.0002) \]

associative learner

\[ r = 0.71 \ (p < 0.01) \]

predictions without further fitting

\[ r = 0.92 \ (p < 0.006) \]

\[ r = -0.23 \ (p > 0.65) \]
MORE MODELS

**familiarisation**
- frequency learner: \( \text{freq}(\text{shape}), \text{freq}(\downarrow), \text{freq}(\text{shape}) \)...
- joint frequency learner: \( \text{freq}(\text{shape} \downarrow), \text{freq}(\downarrow \text{shape}), \text{freq}(\text{shape} \text{shape}) \)...
- conditional probability learner: \( \text{freq}(\text{shape} \downarrow) \div \text{freq}(\text{shape}), \text{freq}(\text{shape} \downarrow \text{shape}) \div \text{freq}(\downarrow) \)...
- associative learner: \( \text{Prob}(\text{pair-wise correlations | familiarization scenes}) \)
- Bayesian chunk learner: \( \text{Prob}(\text{inventory of independent chunks | familiarization scenes}) \)

**test**
- \( \sum \text{for all shapes present } \text{freq}(\text{shape}) \)
- \( \sum \text{for all shape-pairs present } \text{freq}(\text{shape}_1, \text{shape}_2) \)
- \( \sum \text{for all shape-pairs present } \frac{\text{freq}(\text{shape}_1, \text{shape}_2)}{\text{freq}(\text{shape}_1)} \)
- \( \text{Prob}(\text{test scene | pair-wise correlations}) \)
- \( \text{Prob}(\text{test scene | inventory of independent chunks}) \)

**test performance**
- **baseline**
  - basic: \( \text{% correct} \)
- **frequency-balanced**
  - basic: \( \text{% correct} \)
- **triplet**
  - basic embedded: \( \text{% correct} \)
- **quadruple**
  - basic quad pair: \( \text{% correct} \)
  - embedded pair triplet: \( \text{% correct} \)

*Orbán & al, 2008*
THE LEARNING “CURVE”
THE LEARNING “CURVE”
THE LEARNING “CURVE”

inventory:

test:
illusory quad vs mixture quad
THE LEARNING “CURVE”

inventory:

test:

illusory quad vs mixture quad

true pair vs illusory embedded pair
THE LEARNING "CURVE"

inventory:

short training

test:

illusory quad vs mixture quad

true pair vs illusory embedded pair
THE LEARNING “CURVE”

inventory:

short training

inventory:

test:

illusory quad vs mixture quad

true pair vs illusory embedded pair

Probabilistic models of learning and memory — Structural learning

CEU, Budapest, 22-26 June 2009
http://www.eng.cam.ac.uk/~m.lengyel
THE LEARNING “CURVE”

inventory:

short training

test:

illusory quad vs mixture quad
true pair vs illusory embedded pair
THE LEARNING “CURVE”

inventory:

test:

<table>
<thead>
<tr>
<th>illusory quad</th>
<th>mixture quad</th>
</tr>
</thead>
<tbody>
<tr>
<td>true pair</td>
<td>illusory embedded pair</td>
</tr>
</tbody>
</table>

short training

long training

http://www.eng.cam.ac.uk/~m.lengyel

20
THE LEARNING "CURVE"

inventory:

short training

long training

inventory:

short training

illusory quad vs mixture quad

true pair vs illusory embedded pair

true pair vs illusory embedded pair
THE LEARNING “CURVE”

inventory:

short training

long training

test:

illusory quad vs mixture quad

true pair vs illusory embedded pair

Performance (% correct)

Pairs Quads

experiment simulation

Pairs Quads

http://www.eng.cam.ac.uk/~m.lengyel

Probabilistic models of learning and memory — Structural learning

CEU, Budapest, 22-26 June 2009
GOING UP, UP, UP ...


GOING UP, UP, UP ...

data $x$
GOING UP, UP, UP ...
GOING UP, UP, UP ...

\[ P(x|\theta) = \sum_y P(x|y, \theta) P(y|\theta) \]

\[ \hat{\theta} = \arg\max_{\theta} P(x|\theta) \]
GOING UP, UP, UP ...

structure $\mathcal{M}$
\[ P(x|\mathcal{M}) = \sum_\theta P(x|\theta, \mathcal{M}) P(\theta|\mathcal{M}) \]
\[ \hat{\mathcal{M}} = \arg\max_\mathcal{M} P(x|\mathcal{M}) \]

parameters $\theta$
\[ P(x|\theta) = \sum_y P(x|y, \theta) P(y|\theta) \]
\[ \hat{\theta} = \arg\max_\theta P(x|\theta) \]

latent variables $y$
data $x$
GOING UP, UP, UP …

why constrain ourselves to one model form?

structure $\mathcal{M}$

parameters $\theta$

latent variables $y$

data $x$

$$P(x|\mathcal{M}) = \sum_{\theta} P(x|\theta, \mathcal{M}) P(\theta|\mathcal{M})$$

$$\hat{\mathcal{M}} = \arg\max_{\mathcal{M}} P(x|\mathcal{M})$$

$$P(x|\theta) = \sum_{y} P(x|y, \theta) P(y|\theta)$$

$$\hat{\theta} = \arg\max_{\theta} P(x|\theta)$$
GOING UP, UP, UP ...

why constrain ourselves to one model form?

\[ P(x|\mathcal{F}) = \sum_{\mathcal{M}} P(x|\mathcal{M}, \mathcal{F}) P(\mathcal{M}|\mathcal{F}) \quad \rightarrow \quad \hat{\mathcal{F}} = \arg \max_{\mathcal{F}} P(x|\mathcal{F}) \]

\[ P(x|\mathcal{M}) = \sum_{\theta} P(x|\theta, \mathcal{M}) P(\theta|\mathcal{M}) \quad \rightarrow \quad \hat{\mathcal{M}} = \arg \max_{\mathcal{M}} P(x|\mathcal{M}) \]

\[ P(x|\theta) = \sum_{y} P(x|y, \theta) P(y|\theta) \quad \rightarrow \quad \hat{\theta} = \arg \max_{\theta} P(x|\theta) \]
GOING UP, UP, UP …

why constrain ourselves to one model form?

\[
\begin{align*}
P(x|\mathcal{F}) &= \sum_{\mathcal{M}} P(x|M, \mathcal{F}) P(M|\mathcal{F}) & \hat{\mathcal{F}} = \arg\max_{\mathcal{F}} P(x|\mathcal{F}) \\
P(x|M) &= \sum_{\theta} P(x|\theta) & \hat{\theta} = \arg\max_{\theta} P(x|\theta) \\
P(x|\theta) &= \sum_{y} P(x|y, \theta) P(y|\theta) & \text{number of terms: exponential in number of latent variables} \\
\end{align*}
\]

number of possibilities: exponential in number of parameters
GOING UP, UP, UP ...

why constrain ourselves to one model form?

\[ P(x|\mathcal{F}) = \sum_{\mathcal{M}} P(x|\mathcal{M}) \]

\[ P(x|\mathcal{M}) = \sum_{\theta} P(x|\theta, \mathcal{M}) P(\theta|\mathcal{M}) \]

\[ P(x|\theta) = \sum_{y} P(x|y, \theta) P(y|\theta) \]

number of terms: exponential in number of parameters

number of possibilities: exponential in number of model structures

\[ \hat{\mathcal{M}} = \arg\max_{\mathcal{M}} P(x|\mathcal{M}) \]

\[ \hat{\theta} = \arg\max_{\theta} P(x|\theta) \]
why constrain ourselves to one model form?

GOING UP, UP, UP …

number of terms: exponential in number of model structures

number of possibilities: exponential in number of forms

form $F$  
structure $M$  
parameters $\theta$  
latent variables $y$  
data $x$

$P(x|F) = \sum_M P(x|M, F) P(M|F)$  
$\hat{F} = \text{argmax}_F P(x|F)$

$P(x|M) = \sum_{\theta} P(x|\theta, M) P(\theta|M)$  
$\hat{M} = \text{argmax}_M P(x|M)$

$P(x|\theta) = \sum_y P(x|y, \theta) P(y|\theta)$  
$\hat{\theta} = \text{argmax}_\theta P(x|\theta)$
GOING UP, UP, UP ...

why constrain ourselves to one model form?

\[
P(x|\mathcal{F}) = \sum_{\mathcal{M}} P(x|\mathcal{M}, \mathcal{F}) P(\mathcal{M}|\mathcal{F}) \rightarrow \hat{\mathcal{F}} = \arg\max_{\mathcal{F}} P(x|\mathcal{F})
\]

\[
P(x|\mathcal{M}) = \sum_{\theta} P(x|\theta, \mathcal{M}) P(\theta|\mathcal{M}) \rightarrow \hat{\mathcal{M}} = \arg\max_{\mathcal{M}} P(x|\mathcal{M})
\]

\[
P(x|\theta) = \sum_{y} P(x|y, \theta) P(y|\theta) \rightarrow \hat{\theta} = \arg\max_{\theta} P(x|\theta)
\]
### A. Structural Forms and Generative Processes

<table>
<thead>
<tr>
<th>Structural Form</th>
<th>Generative Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partition</td>
<td><img src="image1" alt="Partition Diagram" /></td>
</tr>
<tr>
<td>Chain</td>
<td><img src="image2" alt="Chain Diagram" /></td>
</tr>
<tr>
<td>Order</td>
<td><img src="image3" alt="Order Diagram" /></td>
</tr>
<tr>
<td>Ring</td>
<td><img src="image4" alt="Ring Diagram" /></td>
</tr>
<tr>
<td>Hierarchy</td>
<td><img src="image5" alt="Hierarchy Diagram" /></td>
</tr>
<tr>
<td>Tree</td>
<td><img src="image6" alt="Tree Diagram" /></td>
</tr>
<tr>
<td>Grid</td>
<td><img src="image7" alt="Grid Diagram" /></td>
</tr>
<tr>
<td>Cylinder</td>
<td><img src="image8" alt="Cylinder Diagram" /></td>
</tr>
</tbody>
</table>

Kemp & Tenenbaum, 2008
then search for the structure that best account for the data and the structure.

Inhelder and Piaget recur again and again in formal models across many different literatures. To highlight just one example, Inhelder and Piaget suggested that they are useful for describing the world, and that their models are methods that discover six different kinds of structures given a matrix of binary features.

We take a probabilistic approach, and if \( k \) is small:

\[
\text{p} = \frac{1}{\text{h}_{20841}} \cdot \frac{\text{F}}{\text{s}_{0850}} \cdot \frac{\text{F}}{\text{s}_{11008}} \cdot \frac{\text{F}}{\text{F}_{20862}} \cdot \frac{\text{F}}{\text{F}_{11008}} \cdot \frac{\text{F}}{\text{F}_{20862}}.
\]

Kemp & Tenenbaum, 2008
GRAPH GRAMMARS

A Structural Form Generative process
Partition
Chain
Order
Ring
Hierarchy
Tree
Grid
Cylinder

Chai & Grimes, 2008
Kemp & Tenenbaum, 2008
GRAPH GRAMMARS

A

<table>
<thead>
<tr>
<th>Structural Form</th>
<th>Generative process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partition</td>
<td>![Diagram of a partition]</td>
</tr>
<tr>
<td>Chain</td>
<td>![Diagram of a chain]</td>
</tr>
<tr>
<td>Order</td>
<td>![Diagram of an order]</td>
</tr>
<tr>
<td>Ring</td>
<td>![Diagram of a ring]</td>
</tr>
<tr>
<td>Hierarchy</td>
<td>![Diagram of a hierarchy]</td>
</tr>
<tr>
<td>Tree</td>
<td>![Diagram of a tree]</td>
</tr>
<tr>
<td>Grid</td>
<td>![Diagram of a grid]</td>
</tr>
<tr>
<td>Cylinder</td>
<td>![Diagram of a cylinder]</td>
</tr>
</tbody>
</table>

Kemp & Tenenbaum, 2008
Probabilistic models of learning and memory — Structural learning

DISCOVERING STRUCTURAL FORM

A biological features

B supreme court decisions

C similarity judgements

D pixel values

E geographical distances

Kemp & Tenenbaum, 2008

Probabilistic models of learning and memory — Structural learning

CEU, Budapest, 22-26 June 2009

http://www.eng.cam.ac.uk/~m.lengyel

23
where the entities are faces that vary along two dimensions: described by Newton. Next, we analyzed a similarity dataset structure for these data and corresponds to the color circle observed and the covariance of the data. As long as both of 14 pure-wavelength hues (38). The ring in Fig. 3 containing human judgments of the similarity between all pairs only two components that depend on $S$ in data. The best chain (Fig. 3) Consistent with the unidimensional hypothesis, our model iden-

spaces (36) and sets of clusters (37) have also been proposed. Some political scientists (35) have argued resolution (e.g., mammals, primates, rodents, birds, insects, and other carnivorous land mammals; the songbirds (robin, finch), flying birds categories; and the flying insects (butterfly, bee) and walking insects (ant, cockroach) form distinct subcategories. More information about these simu-

relations can be found in.

LEARNING STRUCTURAL FORM

Probabilistic models of learning and memory — Structural learning

CEU, Budapest, 22-26 June 2009

http://www.eng.cam.ac.uk/~m.lengyel
REFERENCES


Ghahramani (unpubl) http://learning.eng.cam.ac.uk/zoubin/ml06/lect11.pdf
