Feature learning as nonparametric Bayesian inference

By Joe Austerweil and Tom Griffiths
Department of Psychology, UC Berkeley

http://cocosci.berkeley.edu/
What are the features?
Features

• Objects encoded using features.
• Features are the elementary primitives in cognitive models.
• In many cases, the features are not obvious.
• The appropriate feature representation of an object is context-dependent.
The properties of the single stimulus cannot be specified except in relation to the properties of the sets within which it exists.

Wendell Garner
The properties of the single stimulus cannot be specified except in relation to the properties of the sets within which it exists.

Wendell Garner

Thus, representations are flexible and essential to understanding how stimuli are processed.
General computational framework

• Idea: Forming representations is an inductive problem (and thus context dependent)

• Bayesian inference gives us a rational solution to this problem

• Challenge: How do you form a set of possible representations to choose from?

• Flexible hypothesis spaces from nonparametric Bayesian statistics
What is the computational problem?

• Input:
  • Set of raw sensory data of images $\mathbf{x}_i$

• Output:
  • A set of features to represent the images
  • A binary vector of feature assignments for each image ($\mathbf{z}_i$)
  • How that feature is instantiated in images
  • Arbitrarily deep, sparse representations
Learning feature representations

Austerweil & Griffiths (2008)
Learning feature representations

• Goal: form the “best” representation for a set of observed objects, $X$
Learning feature representations

- Goal: form the “best” representation for a set of observed objects, $X$

$$\hat{Z} = \arg \max_Z P(Z|X)$$

Austerweil & Griffiths (2008)
Learning feature representations

• Goal: form the “best” representation for a set of observed objects, $X$

$$\hat{Z} = \arg \max_Z P(Z|X) = \arg \max_Z P(X|Z)P(Z)$$

Austerweil & Griffiths (2008)
Learning feature representations

- Goal: form the “best” representation for a set of observed objects, $X$

$$\hat{Z} = \arg \max_Z P(Z|X) = \arg \max_Z P(X|Z)P(Z)$$

- $P(X|Z)$ is the probability of the objects given features
- Recreates objects given feature matrices
- Accounts for noise in input

Austerweil & Griffiths (2008)
Learning feature representations

- Goal: form the “best” representation for a set of observed objects, $X$

$$\hat{Z} = \arg \max_Z P(Z|X) = \arg \max_Z P(X|Z)P(Z)$$

- $P(X|Z)$ is the probability of the objects given features
  - Recreates objects given feature matrices
  - Accounts for noise in input

- $P(Z)$ is our prior on feature learning matrices
  - Expectations about structure of feature matrices
  - Nonparametric: assumes infinitely many features, but prefers representations with fewer features
  
Austerweil & Griffiths (2008)
Correlated

Independent

Inferred Features

Austerweil & Griffiths (2009)
Feature learning with transforms

Previous feature learning work focuses on features that occur identically across different presentations.

Features can be \textit{transformed} across presentations (Palmer, 1983).
Extending the model to include transforms

Jointly infer the features and their object-specific transformations.

\[ r = 0^\circ, r = 30^\circ, r = 60^\circ, r = 90^\circ \]
Feature learning with transforms

Two object sets where vertical bars are translated either together (unitized) or independently (separable)

People use the set of objects they observe to decide which representation is appropriate.

The smallest representation that can encode the observed objects is used.
Feature learning with transforms

Two object sets where vertical bars are translated either together (unitized) or independently (separable)

People use the set of objects they observe to decide which representation is appropriate.

The smallest representation that can encode the observed objects is used.

Unitized

Feature learning with transforms

Two object sets where vertical bars are translated either together (unitized) or independently (separable).

People use the set of objects they observe to decide which representation is appropriate.

The smallest representation that can encode the observed objects is used.
Feature learning with transforms

Human Results

- Seen both
- Seen Unit
- Seen Sep
- New Unit
- New Sep
- 1 Bar
- Unit + 1 Bar
- 3 Sep Bars
- Diag

Test Image

Human Rating

- Unitized (Unit)
- Separate (Sep)
Feature learning with transforms

Human Results

<table>
<thead>
<tr>
<th>Seen both</th>
<th>Seen Unit</th>
<th>Seen Sep</th>
<th>New Unit</th>
<th>New Sep</th>
<th>1 Bar</th>
<th>Unit + 1 Bar</th>
<th>3 Sep Bars</th>
<th>Diag</th>
</tr>
</thead>
</table>

Test Image

Human Rating

0 2 4 6
Feature learning with transforms

Human Results

Test Image

Human Rating

Seen both  Seen Unit  Seen Sep  New Unit  New Sep  1 Bar  Unit + 1 Bar  3 Sep Bars  Diag

Unitized (Unit)  Separate (Sep)
Feature learning with transforms

Human Results

- Seen both
- Seen Unit
- Seen Sep
- New Unit
- New Sep
- 1 Bar
- Unit + 1 Bar
- 3 Sep Bars
- Diag

Model Predictions

- Seen Both
- Seen Unit
- Seen Sep
- New Unit
- New Sep
- 1 Bar
- Unit + 1 Bar
- 3 Sep Bars
- Diag
Feature learning with transforms
Feature learning with transforms

Human Results

Model Predictions

Test Image

Human Rating

Model Activation
Feature learning with transforms
Feature learning with transforms

Are these two shapes the same?
Feature learning with transforms

Are these two shapes the same?

Should all transforms be included?

Square or diamond?
Feature learning with transforms

Are these two shapes the same?

Should all transforms be included?

Square or diamond?

Hypothesis: people infer the set of transformations allowed for a given shape.
Feature learning with transforms

Contextual effects on allowable transforms

Rotation set
Feature learning with transforms

Contextual effects on allowable transforms

Rotation set

or

?
Feature learning with transforms

Contextual effects on allowable transforms

Rotation set

Size set

or

or

?
Feature learning with transforms

Human Responses

Test Image

Model Predictions

Model Activation
Feature learning with transforms

Human Responses

Test Image

Model Predictions

Model Activation

Human Rating

0 2 4 6

Seen Both Seen Rot Seen Size New Rot New Size

Rotation (Rot) Size
Feature learning with transforms

**Human Responses**

<table>
<thead>
<tr>
<th>Seen Both</th>
<th>Seen Rot</th>
<th>Seen Size</th>
<th>New Rot</th>
<th>New Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5 ± 0.5</td>
<td>4.2 ± 0.6</td>
<td>4.1 ± 0.7</td>
<td>4.0 ± 0.8</td>
<td>4.3 ± 0.4</td>
</tr>
</tbody>
</table>

**Model Predictions**

- **Rotation (Rot)**
- **Size**

**Test Image**

- "Seen Both"
- "Seen Rot"
- "Seen Size"
- "New Rot"
- "New Size"
Conclusions

• Representations are flexible and this flexibility is essential for understanding human cognition.

• Nonparametric Bayesian models can capture the flexibility of human representation learning.

• Infinite number of features, but prefers fewer

• People infer features transformed relative to its object.

  • When ambiguous, people use the smallest feature representation that can encode the observed objects.

• The set of possible transformations depends on the observed objects.
Acknowledgements

• Tania Lombrozo
• Karen Schloss
• Stephen Palmer
• Rob Goldstone
• RAs: David Belford, Brian Tang, Shubin Li, Ingrid Liu, Julia Ying
• CoCoSci, Concepts and Cognition Lab
• You!