Cognitive science for machine learning 2: Empirical Methods

Nick Chater
OVERVIEW

1. OBSERVATIONS AND EXPERIMENTS

2. THE SPECTRUM OF EXPERIMENTAL METHODS

3. CAN YOU PLAY 20 QUESTIONS WITH NATURE AND WIN?
OBSERVATIONS AND EXPERIMENTS
THE CONTRAST

Observations

• Observe human performance in some interesting scenario

• May merely use verbal reports (introspection)

• Or, more usually, turn reports into a task (cf. Signal detection theory, Felix, next session)

• But the most interesting observations are often very general and qualitative

Experiments

• Manipulate one or more independent variables

• Measure one or more independent variables
  – E.g., responses, reaction times (Felix, next session)
  – Body or eye movements
  – Activities of individual neurons
  – Pupil dilation
  – Galvanic skin response
  – EEG waves
  – Cerebral blood flow (fMRI)

OBSERVATIONS, EVEN WHEN INFORMAL, ARE DATA, TOO
Example: Human category learning

- The observation: people can learn to classify the same objects along different dimensions, very flexibly
- But traditional instance-based psychological models, based on a fixed similarity metric don’t...

Heller K, Sanborn A N, & Chater N (2009) *NIPS*
CATEGORIES ARE DEFINED BY DIFFERENT *TYPES OF DIMENSION*
(e.g., object vs substance, Linda Smith et al)

For objects: Shape matters, colour doesn’t
BUT FOR SUBSTANCES...

Shape irrelevant; colour, texture, movement matters
Generalization for just two instances can be powerful---but puzzling for conventional learning accounts.
A Bayesian approach to learning this category structure
Linda Smith (Psychological Review, 1989) showed that children gradually learn to categories more locally, and less globally...
...AND CONDUCT NEW EXPERIMENTS

Sanborn, Heller & Chater, under review
AND QUALITATIVE OBSERVATIONS THAT PEOPLE ROUTINELY...

• Infer causality

• Infer agency

• Infer communicative intent
Inferring causality from observation

Cf. Pearl (2000):

\[ a = a(F,m) = F/m \]

Not \[ F = F(m,a) = ma \]
Inferring agency from observation

From Heider & Simmel, 1944:
Inferring “theory of mind” (with, or without?) powerful innate constraints)
Inferring communicative intentions from observation
Common knowledge is crucial
Communicative, rather than mere signalling, requires common knowledge of joint intentions; and “team reasoning” (Bacharach, Sugden). How might this be learned???
THE STUDY OF VISION IS HEAVILY BASED ON OBSERVATIONS, NOT MERELY EXPERIMENTS

• Kanizsa triangles

• What is the origin of the lightness changes on the edges of the ‘virtual’ triangle?
A 3D variation...
Note the difference...
beginnings of an experiment

Pietro Guardini & Luciano Gamberini: and they have a lovely moving 3D version
OBSERVATIONS MAY SUGGEST GENERAL PRINCIPLES – E.G., FAVOUR THE SIMPLEST EXPLANATION

Find simple abstract patterns... e.g., postulating a square needs 3 parameters; simpler than 7 parameters for accounting for 'cuts' in circles separately.
IF SO, CAN ONLY SHARE INFORMATION WITHIN A SINGLE OBJECT

• Figure-ground separation is different in the two halves
IF SO, CAN ONLY SHARE INFORMATION WITHIN A SINGLE OBJECT

• 1 vs 2 codes for parallel curves
• Such observations reveal quite a lot about visual structure
• Relation to image processing by machine?
WE CAN ALSO OBSERVE, IN A LOT OF DETAIL, THE PATTERNS IN LANGUAGE

• Methodology in linguistics
  – Thesystematisation of intuitions about well formedness

  *John Mary likes
  John likes Mary

  pif vs *fpi

• Or semantics (e.g., the binding contraints):

  • John<sub>j</sub> saw himself<sub>j</sub>
  • John<sub>j</sub> saw him<sub>k</sub>
  • He<sub>i</sub> saw John<sub>j</sub>
  • John<sub>j</sub> saw a picture of Mary and John<sub>k</sub>
...or observation may involve displaying the structure in linguistic input.

DISTRIBUTIONAL ANALYSIS TO RECONSTRUCT SYNTACTIC CATEGORIES (REDINGTON, CHATER & FINCH, 1998)
OR APPLY BAG OF WORDS TOPICS MODELS TO FIND SEMANTIC CATEGORIES

<table>
<thead>
<tr>
<th>217</th>
<th>INSECT</th>
<th>MYB</th>
<th>PHEROMONE</th>
<th>LENS</th>
<th>LARVAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>274</td>
<td>SPECIES</td>
<td>PHYLOGENETIC</td>
<td>EVOLUTION</td>
<td>EVOLUTIONARY</td>
<td>SEQUENCES</td>
</tr>
<tr>
<td>126</td>
<td>GENE</td>
<td>VECTOR</td>
<td>VECTORS</td>
<td>EXPRESSION</td>
<td>TRANSFER</td>
</tr>
<tr>
<td>63</td>
<td>STRUCTURE</td>
<td>ANGSTROM</td>
<td>CRYSTAL</td>
<td>RESIDUES</td>
<td>STRUCTURES</td>
</tr>
<tr>
<td>200</td>
<td>FOLDING</td>
<td>NATIVE</td>
<td>PROTEIN</td>
<td>STATE</td>
<td>ENERGY</td>
</tr>
<tr>
<td>209</td>
<td>NUCLEAR</td>
<td>NUCLEUS</td>
<td>LOCALIZATION</td>
<td>CYTOPLASM</td>
<td>EXPORT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>42</th>
<th>NEURAL</th>
<th>DEVELOPMENT</th>
<th>DORSAL</th>
<th>EMBRYOS</th>
<th>VENTRAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>SPECIES</td>
<td>GLOBAL</td>
<td>CLIMATE</td>
<td>CO2</td>
<td>WATER</td>
</tr>
<tr>
<td>280</td>
<td>SPECIES</td>
<td>SELECTION</td>
<td>EVOLUTION</td>
<td>GENETIC</td>
<td>POPULATIONS</td>
</tr>
<tr>
<td>15</td>
<td>CHROMOSOME</td>
<td>REGION</td>
<td>CHROMOSOMES</td>
<td>KB</td>
<td>MAP</td>
</tr>
<tr>
<td>64</td>
<td>CELLS</td>
<td>CELL</td>
<td>ANTIGEN</td>
<td>LYMPHOCYTES</td>
<td>CD4</td>
</tr>
<tr>
<td>102</td>
<td>TUMOR</td>
<td>CANCER</td>
<td>TUMORS</td>
<td>HUMAN</td>
<td>CELLS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>112</th>
<th>HOST</th>
<th>BACTERIAL</th>
<th>BACTERIA</th>
<th>STRAINS</th>
<th>SALMONELLA</th>
</tr>
</thead>
<tbody>
<tr>
<td>210</td>
<td>SYNAPTIC</td>
<td>NEURONS</td>
<td>POSTSYNAPTIC</td>
<td>HIPPOCAMPAL</td>
<td>SYNAPSES</td>
</tr>
<tr>
<td>201</td>
<td>RESISTANCE</td>
<td>RESISTANT</td>
<td>DRUG</td>
<td>SENSITIVE</td>
<td>CURRENTS</td>
</tr>
<tr>
<td>165</td>
<td>CHANNEL</td>
<td>CHANNELS</td>
<td>VOLTAGE</td>
<td>CURRENTS</td>
<td>LEAVES</td>
</tr>
<tr>
<td>142</td>
<td>PLANTS</td>
<td>PLANT</td>
<td>ARABIDOPSIS</td>
<td>TOBACCO</td>
<td>TASK</td>
</tr>
<tr>
<td>222</td>
<td>CORTEX</td>
<td>BRAIN</td>
<td>SUBJECTS</td>
<td>AREAS</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>39</th>
<th>THEORY</th>
<th>TIME</th>
<th>SPACE</th>
<th>GIVEN</th>
<th>PROBLEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>105</td>
<td>HAIR</td>
<td>MECHANICAL</td>
<td>IMB</td>
<td>SENSORY</td>
<td>EAR</td>
</tr>
<tr>
<td>221</td>
<td>LARGE</td>
<td>SCALE</td>
<td>DENSITY</td>
<td>OBSERVED</td>
<td>OBSERVATIONS</td>
</tr>
<tr>
<td>270</td>
<td>TIME</td>
<td>SPECTROSCOPY</td>
<td>NMR</td>
<td>SPECTRA</td>
<td>TRANSFER</td>
</tr>
<tr>
<td>55</td>
<td>FORCE</td>
<td>SURFACE</td>
<td>MOLECULES</td>
<td>SOLUTION</td>
<td>SURFACES</td>
</tr>
<tr>
<td>114</td>
<td>POPULATION</td>
<td>POPULATIONS</td>
<td>GENETIC</td>
<td>DIVERSITY</td>
<td>ISOLATES</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>109</th>
<th>RESEARCH</th>
<th>NEW</th>
<th>INFORMATION</th>
<th>UNDERSTANDING</th>
<th>PAPER</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td>AGE</td>
<td>OLD</td>
<td>AGING</td>
<td>LIFE</td>
<td>YOUNG</td>
</tr>
</tbody>
</table>

SUCH ANALYSIS OF THE STRUCTURE OF THE ENVIRONMENT LOOKS A LOT LIKE MACHINE LEARNING

• And non-rigid boundary between
  – characterising the information in the input vs
  – providing a computational model of cognition

• Link with empirical data
  – E.g., Griffiths, Steyvers, Tenenbaum, 2007, Psych Review, on word associations
OBSERVATIONS, WHETHER EVERYDAY OR CLEVERLY CONSTRUCTED, ARE VITAL CLUES

• Vision (interpretations of visual images, illusions)
• Language (intuitions underlying linguistics, semantics)
• Common-sense knowledge and reasoning

• But observations are not enough---in particular to establish causal relationships between variables, it is often vital to conduct experiments
THE SPECTRUM OF EXPERIMENTAL METHODS
Simple psychophysics - detection

- On half the trials, there really is a faintly brighter square
- 2AFC response
- Manipulate, e.g., whether/how many blobs are cut by the square/size of blobs
- Measure detection rate (signal detection theory)
Simple psychophysics - detection

- We might manipulate display duration
- Response deadline
- Density of dots
- Size of the stimulus
- Side of the square...

- But we can also consider a richer dependent variable, e.g., RTs
FROM RESPONSES TO REACTION TIMES

• Helmholtz measure of speed of nervous conduction (roughly 27m/s)
FROM RESPONSES TO REACTION TIMES

• Helmholtz measure of speed of nervous conduction (roughly 27m/s)
FROM RESPONSES TO REACTION TIMES

• Helmholtz measure of speed of nervous conduction (roughly 27m/s)
FROM RESPONSES TO REACTION TIMES

- Helmholtz measure of speed of nervous conduction (roughly 27m/s)
FROM RESPONSES TO REACTION TIMES

- Helmholtz measure of speed of nervous conduction (roughly 27m/s)
FROM RESPONSES TO REACTION TIMES

• Helmholtz measure of speed of nervous conduction (roughly 27m/s)
BUT OFTEN IT IS VALUABLE TO MEASURE INCIDENTAL VARIABLES

– e.g., eye-tracking reveals incremental interpretation in reading (e.g., Traxler, Pickering et al) – slow down on the word that doesn’t fit...
AND WE MAY MEASURE THE BRAIN, NOT JUST THE EYE...

- Event-related potentials (ERP) have characteristic patterns for syntactic and semantic anomaly (e.g., Kutas)
AND WE MAY WONDER WHERE IN THE BRAIN (PET, FMRI)

Specific regions appear associated with fear, surprise, different locations in the visual field, motor planning, social emotions, and so on.

So a rich new set of dependent variables... though always a danger of their interpretation being circular.
Imaging technologies provide another role for machine learning

- What can we really infer?
- Potentially horrible problems of multiple comparisons

Cause celebre: fMRI of social emotions in a dead salmon (Craig Bennett, 2009)
MACHINE LEARNING CAN HELP

• But the results may be rather demoralizing...???

CAN YOU PLAY 20 QUESTIONS
WITH NATURE AND WIN?

Alan Newell 1927-1992
EXPERIMENTAL PSYCHOLOGY AND THE SEARCH FOR BINARY DIVISIONS

• Experiments can at best answer questions posed to choose between broad classes of option

  – Face processing system vs general visual perception?
  – Semantic vs episodic memory
  – Short vs long term memory
  – Hot vs cold cognitive processes
  – ...
  – ...
  – ...
  – ...
BUT EXPERIMENTAL PSYCHOLOGY CAN NEVER SUCCEED ALONE

• Newell (1973) argued that most such debates are inconclusive

• And even if conclusive 20 bits of information is not enough to specify a theory of the mind

• Observations/intuitions are vastly richer, and provide very rich constraints (as we have seen)

• And perhaps the biggest constraint of all is functional: good reverse engineering must be good engineering!
BUT WE NEED OUR 20 QUESTIONS NONETHELESS

• ...unless there is one solution to the problem of intelligence

• Cognitive science and machine learning need to ask the right questions, splitting along genuinely critical issues
  – Discriminative vs model-based perception???
  – Modular vs non-modular processes???
  – Exemplar vs rule-based process???
  – Symbolic vs connectionist cognition???

FINDING THESE QUESTIONS IS A CHALLENGE FOR THE REST OF THIS WORKSHOP, AND BEYOND!