An Introduction to Learning

Lecture 13/13

Todd M. Gureckis
Department of Psychology
New York University
todd.gureckis@nyu.edu
Agenda for Today

- Final
- Midterm grades
- Questions from last time
- Closing arguments
Why did the fully connected networks fail?
Midterm Grades (so far)

How Many People Answered Each Question?

<table>
<thead>
<tr>
<th>Question</th>
<th>Number of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Define Learning</td>
<td>5</td>
</tr>
<tr>
<td>2. Lashley</td>
<td>12</td>
</tr>
<tr>
<td>3. Neural mechanisms</td>
<td>9</td>
</tr>
<tr>
<td>4. Tolman</td>
<td>10</td>
</tr>
<tr>
<td>5. Perceptual Learning</td>
<td>15</td>
</tr>
<tr>
<td>6. Rescorla-Wagner</td>
<td>11</td>
</tr>
<tr>
<td>7. TD</td>
<td>11</td>
</tr>
<tr>
<td>8. Prediction errors</td>
<td>6</td>
</tr>
<tr>
<td>9. CSUS</td>
<td>8</td>
</tr>
<tr>
<td>10. Latent Cause</td>
<td>4</td>
</tr>
<tr>
<td>11. Education</td>
<td>9</td>
</tr>
<tr>
<td>12. Multiple systems</td>
<td>8</td>
</tr>
<tr>
<td>R. Learning/your research</td>
<td>18</td>
</tr>
</tbody>
</table>
Key Principals for the Semester

- Learning and memory are closely related and intertwined states of information processing
- Major insights about learning and memory have come from studies of the brain
- The concept of multiple memory systems unifies the study of learning and memory
- The underlying bases of learning and memory are the same in humans and animals
- Our theoretical approaches to studying learning are always closely tied to technological advances that are unfolding in general society (e.g., today - machine learning)
The Need for Biases in Learning Generalizations

Tom M. Mitchell
The Generalization Problem
Given:

1. Language of instances.
2. Language of generalizations.
3. Matching predicate for matching generalizations to instances.
4. Sets of positive and negative training instances.

Determine:

⇒ Generalization(s) consistent with the training instances.
In Machine Learning it might be desirable to have an unbiased generalizer (i.e., this would be free of “arbitrary constraints”)

Where do biases come from?

The language of generalizations (i.e., hypothesis space in a Bayesian model)

The method of searching for or finding generalizations within some space.
The language of generalizations
(i.e., hypothesis space in a Bayesian model)

- You can only learn generalizations that the system already knows how to represent (turns into a Fodorian problem... you can’t learn anything if you didn’t already “know” it)
The method of searching for or finding generalizations within some space.

- An unbiased search procedure looks through possible solutions to find a potential generalization that includes all “positive” instances and none of the “negative” ones.

- However, it must do this in an unbiased way (so that it is equally likely to come up with any of the possible generalizations that are consistent)

- An unbiased classifier would predict the membership of some new observation as a member of the set if all possible generalization agree it is in the set, and not if all agree it is out of the set, or could response probabilistically.
3. The Futility of Removing Biases

Although removing all biases from a generalization system may seem to be a desirable goal, in fact the result is nearly useless. An unbiased learning system’s ability to classify new instances is no better than if it simply stored all the training instances and performed a lookup when asked to classify a subsequent instance.

(the only set of generalization they all agree on will be a table-lookup)
“[All] The power of a generalization system follows directly from its biases-- the decision based on criterion other than consistency with the training instances”

**Useful Classes of Biases**

- **Factual knowledge of the domain.**

- **Intended use of the learned generalizations.**

- **Knowledge about the source of training data.**

- **Bias toward simplicity and generality.**

- **Analogy with previously learned generalizations.**
**Factual knowledge of the domain.**
(i.e., prior knowledge/or the evolutionary prior of the system)

- **CS-US-CR compatibility:** Different USs are more easily conditioned to certain CSs

- Garcia & Koelling (1966) showed that rats learned to avoid drinking sugar water when injected with drug to make them sick, but not when shocked afterwards (easier to learn food->sick than food->shock).

  - Also depends on the animal species: pigeons more easily associate color with illness, rats associate outcomes with flavors

- Evolutionary Adaptive constraints for learning particular forms of association
The concept of “Belongingness”

- Just as certain CS-US relationships are easier to acquire, certain behaviors are easier to condition instrumentally (remember Garcia & Koelling, 1966 finding on conditioning tastes/sickness/shock):
  - Harder to get cats to condition grooming to escape relative to playing with strings (Thorndike)
  - At zoo pigs trained to drop coin in a slot try kicking it rather than dropping with mouth
  - ...and raccoons keep playing with the coins and never dropping them in (Breland & Breland, 1961)
  - and yet instrumental conditioning remains incredibly powerful...

http://www.youtube.com/watch?v=Nc9xq-TVyHI
Intended use of the learned generalizations.

**ABA -> GFG**

(but only for linguistic sounds, not for musical sounds)
Knowledge about the source of training data.

pedagogical assumption (strong vs. weak sampling)

Figure 1: Possible rectangle game scenarios. The top row shows a possible rectangle concept, and two possible pairs of examples that a teacher might choose to communicate to a learner. The bottom row shows possible examples a learner may observe, and two possible guesses about what rectangle the teacher had in mind. The middle column shows better choices than the right column.

Figure 2: Distributions of examples in the teaching task for (a) positive examples and (b) negative examples. Pictured from left to right in each panel are the predictions of strong and weak sampling, the observed human data, and the predictions of pedagogical sampling. For the models, figures display the probability of an example in each block. For human data, the proportion of positive examples in each location is plotted. People strongly preferred to give positive examples in the corners of the rectangle and negative examples near the boundaries, as predicted by pedagogical sampling.

Figure 4: Results from the learning task. Plots show the positions of the teacher’s examples, relative to the rectangles drawn by learners for positive (left) and negative (right) examples. The results show that learners clearly understand that teachers are sampling data pedagogically – positive examples indicate corners of the correct rectangle, and negative examples indicate the boundaries.
Bias toward simplicity and generality.

(b) Marginal likelihood

Learning via Bayes Rule

\[ P(S|D) = \frac{P(D|S)P(S)}{\sum_{S'} P(D|S')P(S')} \]
Analogy with previously learned generalizations.

DIFFERENTIATION

FIG. 7.9. Materials used by Pevtzow and Goldstone (1994). The four objects, A, B, C, and D, were categorized into two groups. When A and B were placed in one group, and C and D were placed in the other, the parts on the right were diagnostic. When A and C were placed in one group, and B and D were placed in the other, then the parts on the bottom were diagnostic.
What are the remaining challenges for learning theory? Is it basically understood well enough now? What are the puzzles?

- Why are there critical periods in learning for some kind of things and not others?
- What are there u-shaped patterns in Cognitive Development and Learning?
- What are the set of inductive biases that human learners exhibit and how they change depending on the current situation?
- Why does the brain fractionate into multiple, specialized systems. Is it a “kludge” or reflection of an ideal solution to the problem of adaptive behavior?
- How do the multiple forms of learning we’ve covered (e.g., instrumental, learning by analogy, observational learning) interact to control behavior?
- Why does motivation play such a powerful role in learning, and what are the factors that influenced motivated learning?
What do I think the field of learning will look like in 50 years?

- (who cares what you think!)

- In contrast to today’s emphasis on isolating the contribution of isolated system to learning, the question will shift more fully to how multiple, interacting brain regions give rise to behavior.

- Learning will be studied in richer contexts where multiple data sources are leveraged to guide inference (owing the machine learning/data mining advances).

- Computational theories will be at the center point for making sense of all this data. The problem in the next 50 year in many areas of science is the amount of data we can collect. We need more powerful ways of thinking about it.

- We will find that many of the basic mechanism identified by Pavlov, Thorndike, Skinner, Tolman, actually are at the root of many real world human behaviors. It’s not just the 2% of the variance in learning behavior (as one might actually suppose sometimes given the disconnect between the quantitative study of learning and practical implications of that work).

- The “framework” of inductive bias will gradually replace the nature/nurture debate by helping to clarify what people mean and laying out better, more reasonable criteria for understanding the interplay of experience and biological structure in adaptive behavior.
Thanks!