Reinforcement and Representation

A review of the behaviorism debates, and their implications for the study of Reinforcement learning

Humanoids Seminar
November 20, 2009
Jon Scholz
First, an experiment

- Can you count how many times the white team passes the ball?
And the point of that?
(What does attention have to do with behaviorist theories??)

- Perception of stimuli depends on the internal state of the observer
- Can you predict who will notice the gorilla?
- I can change this probability by simply changing what I say before the video
- (Radical) behaviorism eschews the notion of internal state
- Really? ... yes!
Behaviorism and learning

- Has had great success at explaining many complicated behaviors in empirical studies
- Has received considerable criticism at explaining human behavior
  - Makes striking assumptions about the learning process
- Inspired the dominant paradigm for artificial agent-based learning (RL)
learning and Robotics

- Humanoid robots need human-type behaviors
- Reinforcement Learning is a promising approach to acquiring behaviors autonomously
- RL owes its original model to behaviorism
  - behaviorist studies haven’t explained uniquely-human behaviors
- Our goal: use radical behaviorism as a foil to explain the fundamental learning problem for humanoid robots
An outline of behaviorism

- A theory that attempts to explain human intelligence as a product of reinforced behaviors

- everything an organism does is a “behavior”

- gets rid of wishy-washy terms like “thoughts”, “feelings”, “knows”, “wants”

- Historically viewed as a reaction to 19th century mentalist psychology (introspection, mental states)

- based on Aristotle’s (and our) intuitive feeling that learning is the association of ideas
How does it work?

1: Classical Conditioning

- Some stimuli innately produce behaviors (US->UR)
  - salivating for good smells, recoiling from pain
- Most stimuli don’t (NS)
- Animals can be trained to associate NS with UR:
  - US+CS->UR
  - CS->UR
- *key stimulus is before response
How does it work?

2: Operant Conditioning

- Skinner adapted classical conditioning to explain behavior modification
  - modeled changes in responses to stimuli presented after the behavior
  - Explains how rats learned to quickly escape those twisted little Skinner boxes
Operant Conditioning: what it looks like

- Behavior modification follows canonical laws based on the type and pattern of stimuli
- Reinforcement
- Punishment
- Extinction
What operant conditioning can do

- Shape any behavior that spontaneously emerges during a reinforcement event
- Rat can be taught to press a lever to obtain food
- Dolphins can be taught to jump through hoops at SeaWorld
- Dogs can be trained to heel
What about learning in people?

- Children can be trained to be polite to grown-ups or study to get good grades
- Skinner was very interested in extending his theory to human behavior
  - Wrote *Verbal Behavior* in 1957 to explain word learning in children
  - Proposed that all human learning could be explained by the laws of operant conditioning
Explaining world learning with behaviorist principles

- Thoughts?
- How about: “This is a pencil”
  - What’s the target stimulus?
  - Where’s the reinforcer?
- Is this a good theory?

That’s okay, skinner has notion of “secondary reinforcers”
A sketch of Verbal Behavior

- A theoretical (not empirical) account of learning
- Applied vocabulary of operant conditioning to word learning (“functional analysis”)
- Objects and events are the controlling stimuli
- Words are the operants

<table>
<thead>
<tr>
<th>Precondition</th>
<th>Verbal Operant</th>
<th>Consequence</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivating Operation</td>
<td>Mand</td>
<td>Directly effective</td>
<td>A child comes into the kitchen where a mother is, and says: &quot;I want milk&quot;. The mother opens the refrigerator and gives the child milk.</td>
</tr>
<tr>
<td>Feature of the physical environment</td>
<td>Tact</td>
<td>Social</td>
<td>A student looks out of the window, turns to his teacher and says: &quot;It is hot today.&quot; The teacher says, &quot;Right!&quot;</td>
</tr>
<tr>
<td>Verbal behavior of another person</td>
<td>Intraverbal</td>
<td>Social</td>
<td>A mother asks her daughter: &quot;What grade did you get in math?&quot; The daughter replies, &quot;An A.&quot; The mother says: &quot;Very good!&quot;</td>
</tr>
<tr>
<td>Verbal behavior of another person</td>
<td>Echoic</td>
<td>Social</td>
<td>A teacher says to a student: &quot;Behavior in German is Verhalten.&quot; The student repeats &quot;Behavior is Verhalten.&quot; The teacher says &quot;Correct.&quot;</td>
</tr>
<tr>
<td>A person's own verbal behavior</td>
<td>Autolitic</td>
<td>Directly effective</td>
<td>A child comes into his parents' bedroom at night and says &quot;I think I am sick.&quot; The mother takes the child and brings him to a hospital.</td>
</tr>
</tbody>
</table>
Our review

- Is this still a good theory?
- How should we handle the pencil problem with this stimulus:
  - “This is a pencil”
    - What is?? Woman? Pencil? Woman holding pencil? Blonde hair?
  - “What is this picture of?”
  - “A pencil”, “An advertisement”, “My neighbor”
Other problems? Chomsky thinks so

- Chomsky wrote a very critical review of Verbal Behavior in 1959, which was widely credited for the decline in popularity of behaviorism and the emergence of cognitive psychology.

-Outlined 3 main issues
  - No empirical work: extrapolation to higher cognition
  - Language is generative and combinatorial (how is a sentence reinforced if it’s never been uttered before?)
  - *Perception of stimulus depends on hidden information (internal state)*
Time for an example

- What would Skinner say to the operant “Dutch” when a subject is presented with a painting?
- response asserted to be “under the control of extremely subtle properties” of the physical object or event
- Suppose instead of “Dutch”, the subject said “clashes with wallpaper”, “I thought you liked abstract work”, never saw it before”, “tilted”, “hanging to low”, “beautiful”, “remember our camping trip last summer”,…
- Skinner must say each is under a control of a different “stimulus property”
  - e.g. if we look at a red chair, we’re under control of “redness”, but if we say “chair” we’re under control of “chairness”
Trouble for the theory?

- Operant conditioning requires that we know which physical event is the stimulus on a given occasion.
- Experimental psychologist faces a dilemma:
  - If we define all events as stimuli, then must concede that some behaviors are not lawful (“this is a pencil” -> ”this is a middle finger”)
  - If we consider only lawful behavior, then most of language is not considered behavior.
so, what’s wrong with that?

- “Stimulus” loses all objectivity
  - Stimuli are no longer part of the outside world, they’re “driven back into the organism”

- Empirical problem too:
  - We cannot predict verbal behavior in terms of the stimuli in the speaker’s environment, since we do not know what the current stimuli are until he responds

- Net effect is to relax definitions of stimulus and response until they’re no more helpful than mentalist terms

\[\text{can’t win unless you cheat...}\]
What does this means about learning??

- Behaviorist core assumption is too strict:
  
  Implies that there's processing going on inside the individual that governs the response to the stimulus, and that it can vary between trials or individuals for physically identical stimuli

- In other words, we need to resort to mentalist terms like "knows, believes, wants, etc" to explain human behavior.
**Summary of argument**

- Association between stimuli and rewards can’t explain the acquisition of optimal behaviors without granting also the agent's ability to disambiguate and deliberate.
- In other words, we need to consider an internal representation.
Relationship to Reinforcement Learning

- RL based on
  - agents coming up with a policy
  - by exploring states
  - and receiving rewards
The RL model

- RL is a sequential decision problem formalized by an MDP
  - MDP: \(<S,A,T,R,\gamma>\)
  - \(S\) = State space
  - \(A\) = Actions
  - \(T\) = Transition model
  - \(R\) = (external) Reward function
  - \(\gamma\) = discount factor

- In RL, agent doesn’t have \(T\) or \(R\)
  - tough, needs to behave right anyway
RL to OC mapping

- Action = Operant
- State = Stimulus
- Reward = Reward
- Model = (general notion of memory and planning)
Solving an MDP in RL

- Basic idea: explore
  - update the expected reward for the transition you just made using the reward you actually got
- Works great if you explore long enough
Variants of the RL model

- Model-free ("primitive"/"direct")
- Model-based (learn R and T, then solve MDP old fashioned way)
- There’s more, ask peng or charles
  - “Relational”
  - “heirarchical”
  - etc...
What sort of stuff can we do?

- Loads of stuff. It’s so general purpose!
Limitations?

- Curse of dimensionality
  - State space is exponential in the number of features (dimensions)
  - To find an optimal policy we have to explore lots of this space
- Thus, bad at solving high level problems (e.g. walking, cleaning, building) when posed with low-level features (joint angles, pixels)
Solutions?

- Function approximation!
  - great, but doesn’t always work and isn’t well understood
  - not powerful enough to bridge the gap for hard problems
    - most we can hope for is a fast enough computer or a smooth enough function to get it with brute force
- Really, we’re better off giving the MDP the right features
  - Object pose
  - End-effector pose
What we’re really doing when we solve an MDP

- Same thing that rats are doing in a skinner box
  - associating good actions to perceived states
    - States are built in to the system from the start
  - Policy learning doesn’t give good representation, only good actions
To solve high level problems:

- We provide the right representation!
- This is the hardest part of making a robot behave properly!
  - Perception problem: have to go from pixels to objects, objects to relations, etc
  - Action problem: have to go from joint currents to pushing, pushing to placing, grasping, assembling, etc...
So what’s wrong with designing high-level representations?

- Sometimes can be hard or impossible
  - E.G. Walking involves well understood physics, but there is no representation that allows a robot to walk like humans
- And sometimes we don’t even know what the right representation would look like
  - Theory of Mind, program authoring, storytelling
Since Chomsky thinks he’s so smart, what does he propose?

- Well, he proposed that the necessary bits are innate.
- But his successors provided more moderate (and helpful) arguments.
  - Learning as theory formation (remember this?)
  - Starting state nativism
- Even proposed specific computational accounts of learning.
Supporting evidence for bias and concept learning in people:

- Can look for evidence of internal bias in a learner
  - universal grammar, face bias, shape bias, ...
- Can Look for evidence that learners build models from data
  - worlds like “google” or “unfriend”
  - what a “bird” is
  - intuitive physics
  - how to print in CCB
Sounds great, but can we do it?

- Hopefully - there’s certainly exciting progress in Cognitive Science (remember the Theory² talk?)
- Peng and I also have some specific ideas
- But really, this talk is more about putting robot learning in perspective using theories of human learning
- As we set our sights on higher level tasks, it seems important to take evidence from earlier solutions: ourselves!
Representing Cause and effect

- Core idea: capture it all in a bayes net
- Edges reflect conditional dependence (conditional independence implied by lack of edge)
Experiment IV: ambiguous condition

- $P(E|A) = 1$
- $P(E|B) = 1$
Can this be automated?

- Same as verbal behavior argument
  - Exploration and reward can’t give us high level representations if we’re only learning a policy