Psychology

Friday, June 26, 2020 1:07 PM

- Algorithms are mainly termed as for prediction or control, which roughly correspond to the following terms used by Psychologists: classical or Pavlovian conditioning and instrumental or operant conditioning
  - The correspondence between prediction algorithms and classical conditioning rests on their ability to predict upcoming stimuli
- Pavlovian (classical) conditioning revolves around connecting new stimuli to innate responses
  - The innate responses are usually called unconditioned responses (URs)
  - The triggering stimuli are usually called unconditioned stimuli (US)
  - The new responses triggered by the predictive stimuli are usually called conditioned responses (CRs)
  - Once they have been learned the triggering stimuli become called conditioned stimuli (CSs)
- Delay Conditioning:
  - A classical conditioning experiment where the CS extend throughout the interstimulus interval (ISI, time between the onset of the CS and the onset of the US)
- Trace Conditioning:
  - A classical conditioning experiment where the US begins after the CS has ended
    - In this setup the time between the CS ending and US beginning is called the trace interval
    - The name comes from the idea from Pavlov that there must be a trace left in the nervous system that persists for some time after the stimulus ends for this conditioning to occur
      - Eligibility traces are like a version of this trace but applied to an animal's own actions, not an external stimulus (the idea of an internal action trace was proposed by Hull)
- Blocking:
  - When an animal fails to learn a CR when a potential CS is presented along with a CS that had previously been used to condition the animal to produce the CR (i.e. adding a second potential stimulus to an already trained stimulus does not result in any learning of the new stimulus)
    - This is important since it refutes the idea that a necessary and sufficient condition for conditioning is that a US frequently follows a CS closely in time
- Higher Order Conditioning:
  - When a previously conditioned CS acts as a US in conditioning another initially neutral stimulus (i.e. a conditioned stimulus is used to create further conditioned responses)
    - The degree of separation between the original stimulus and the current one is used to term this second-order, third-order and so on
    - This applies to classical and instrumental conditioning
- Rescorla-Wagner Model:
  - The idea that an animal only learns when events violate its expectations
    - Note that the expectations don't have to be conscious
  - Equations:
  - $\Delta V_A = \alpha_A \beta_Y \left(R_Y - V_{AX}\right)$
  - $\Delta V_X = \alpha_X \beta_Y \left(R_y - V_{AX}\right)$
    - Where V_x is the "associative strength" of stimulus x, $V_{xy} = V_x + V_y$ is an "aggregate associative strength" associated with the stimulus compound xy (a compound stimulus formed from component stimuli x and y, literally this means that the stimuli are presented together such as a sight and sound appearing simultaneously), α and β are step-size parameters, and R_x denotes the asymptotic level of associative strength that the US x can support
    - The updates are the obvious choice of just adding these to the existing values
    - Larger values of V are intended to lead to stronger or more likely CRs and negative V would

mean that there would be no CR, although the exact details of this mapping aren't specified

Note that this model explains blocking since adding a new stimulus to a stimulus or compound stimulus that is already at or near the asymptotic threshold will result in no or very little change in the associated strengths

This model was important since it showed that blocking could be explained by a mathematical theory without resorting to more complex cognitive theories

In the Notation of this Book:

$\vec{w}_{t+1} = \vec{w}_t + \alpha \delta_t \vec{x}(S_t)$

$\delta_t = R_t - \hat{v}(S_t, \vec{w}_t)$

$\hat{v}(s, \vec{w}) = \vec{w}^T \vec{x}(s)$

Where t refers to the number of a complete trial and not its usual meaning of a single time step, the feature vector $\vec{x}(s)$ is a binary encoding where each element represents whether or not the corresponding stimulus is present, and $\hat{v}(s, \vec{w})$ represents the aggregate associative strength for trial-type s

Since the only meaningful time step in this model is a complete trial, this is called a trail-level model

$\hat{v}(s, \vec{w})$ is called a value estimate in reinforcement learning, but a US prediction in psychology

The prediction error δ_t can be thought of as a measure of surprise

TD Model of Classical Conditioning:

$\vec{w}_{t+1} = \vec{w}_t + \alpha \delta_t \vec{z}_t$

$\delta_t = R_{t+1} + \gamma \hat{v}(S_{t+1}, \vec{w}_t) - \hat{v}(S_t, \vec{w})$

$\vec{z}_t = \gamma \lambda \vec{z}_{t-1} + \vec{x}(S_t)$

Where $\vec{z}_t$ is a vector of eligibility traces (as should be clear from the definition), $\hat{v}(s, \vec{w})$ refers to the aggregate associative strength as defined above, and unlike the Rescorla-Wagner model, t here refers to normal time steps

Since this model uses normal time steps instead of trial level time steps, this is called a real-time conditioning model

This allows it to capture subtleties in the timing of events that the Rescorla-Wagner model cannot (e.g. the neutral stimulus should obviously come before the response for conditioning to occur)

Stimuli representations for this model include the complete serial compound representation, where no temporal generalization occurs for states, the presence representation, where complete temporal generalization occurs for states (a state is represented only as present or not), and the microstimuli representation, where some temporal generalization occurs

Temporal generalization as I use it above refers to whether a state is considered different if it occurs at different times (I think)

The microstimuli representation is probably the best biological analog

Note that this model is a generalization of the Rescorla-Wagner model

The Rescorla-Wagner model is recovered if $\gamma = 0$ albeit with the lingering differences that the meaning of the time step is different and in the TD model the prediction target R is on a one-step lead

This model accounts for higher-order conditioning, which the Rescorla-Wagner model does not

This is due to the use of the TD error, which can form associations without a direct link to the US since R_{t+1} has the same status as $\gamma \hat{v}(S_{t+1}, \vec{w}_t) - \hat{v}(S_t, \vec{w})$

Generally, bootstrapping is related to higher-order conditioning

Note that this is basically semi-gradient TD(λ) with linear function approximation

Serial Compound:

A compound stimulus where the component stimuli occur in a sequence over time

Extinction Trials:

Trials where the associative strength of a CS decreases over time

This is often the case in tests of higher-order conditioning, where the original stimulus loses associative strength since it is no longer being associated with the US

- Features of Reinforcement Learning Algorithms:
  - They are selectional
    - They try alternatives and select among them by comparing their consequences
  - They are associative
    - The alternatives found by selection are associated with particular states to form the agent's policy
- The features above have computational analogs in search and memory
- Shaping:
  - A process of reinforcing a behavior through reinforcing successive approximations of the desired behavior
- An environmental model has two parts, a state-transition part that encodes knowledge about the effect of actions on state changes and a reward part that encodes knowledge about the reward expected for each state or state-action pair
- Cognitive Map:
  - An internal model of an animal's environment (sometimes the term task-space is used)
- Expectancy Theory:
  - The idea that animal's learn cognitive maps through stimulus-stimulus associations, where the occurrence of a stimulus generates an expectation about the stimulus to come next
- Model-free and model-based reinforcement learning methods correspond to the distinction psychologists make between habitual and goal-directed control of learned behavior patterns
- Outcome-Devaluation Experiments:
  - Experiments where after initial conditioning the reward value of an outcome is reduced, including to 0 or a negative value
- With repetition goal-directed actions become habitual
  - One idea to model this is that animals use both model-free and model-based methods, which both propose an action, then the action chosen to execution is that judged to be more trustworthy as determined by measures of confidence maintained throughout learning
    - Early in learning the model-based method is more trustworthy since it can generally learn faster but later in training the model-free method becomes more trustworthy since models are perfect and generally involve shortcuts