Week 4 Video 7

Memory Algorithms

Is future correctness enough?

- Up until this point we've been talking about predicting future correctness

But what if you forget it tomorrow?

- Another way to look at knowledge is – how long will you remember it?

Relevant for all knowledge

- Mostly studied in the context of memory for facts, rather than skills

- How do you say banana in Spanish?
- What is the capital of New York?
- Where are the Islands of Langerhans?

Spacing Effect

- It has long been known that spaced practice (i.e. pausing between studying the same fact) is better than massed practice (i.e. cramming)
- Early adaptive systems implemented this behavior in simple ways (i.e. Leitner, 1972)

ACT-R Memory Equations (Pavlik & Anderson, 2005)

- Memory duration can be understood in terms of memory strength (referred to as activation)

ACT-R Memory Equations (Pavlik & Anderson, 2005)

- Formula for probability of remembering

- $P(m) = \frac{1}{1+e^{\frac{\tau - m}{s}}}$

- Where m = activation strength of current fact
- τ = threshold parameter for how hard it is to remember
- s is noise parameter for how sensitive memory is to changes in activation

- Note logistic function (like PFA)

ACT-R Memory Equations (Pavlik & Anderson, 2005)

- Formula for activation
- $m_n(t_{1\ldots n}) = \ln\left(\sum_{i=1}^{n} t_i^{-d}\right)$
- We have a sequence of n cases where the learner encountered the fact
- Each t_i represents how long ago the learner encountered the fact for the i-th time
- The decay parameter d represents the speed of forgetting under exponential decay

ACT-R Memory Equations (Pavlik & Anderson, 2005)

- Implications
- More practice = better memory
- More time between practices = better memory
  - Most efficient learning comes from dense practice followed by expanding amounts of time in between practices (Pavlik & Anderson, 2008)

MCM (Mozer et al., 2009)

- Postulates that decay speed drops, the more times a fact is encountered
- Functionally complex model where
  - Knowledge strength (and therefore probability of remembering) is a function of the sum of the traces' actual contributions, divided by the product of their potential contributions
  - Power function is estimated as a combination of exponential functions

DASH (Mozer & Lindsay, 2016)

- DASH Extends previous approaches to also include item difficulty and latent student ability
- Can use either MCM or ACT-R as its internal representation of how memory decays over time

Duolingo (Settles & Mercer, 2016)

- Fits regression model to predict both recall and estimated half-life of memory (based on lag time)
- Based on estimate of exponential decay of memory

Duolingo (Settles & Mercer, 2016)

- Uses feature set including
  - Time since word last seen
  - Total number of times student has seen the word
  - Total number of times student has correctly recalled the word
  - Total number of times student has failed to recalled the word
  - Word difficulty

Another area of active development

- Watch this space, approaches rapidly changing
- Recent emerging approaches have not yet gone “head to head” against each other

Next Week

- Relationship Mining