

Optimizing Paired-Associate Learning

Philip I. Pavlik (ppavlik@andrew.cmu.edu)
John R. Anderson (ja+@cmu.edu)

Psychology Department and the Center for the Neural Basis of Cognition, Carnegie Mellon University

• This research tries to answer the question: what is the maximally efficient schedule of practice when one needs to learn a collection of facts?

- Could apply in a number of domains:
 - Math Facts
 - Spelling
 - Language Vocabulary (Native or Foreign)
 - History Facts
 - Geography Facts

• In this work a training program uses a cognitive model to predict before each trial the most efficient item to practice next.

• An interesting consequence of this optimized learning is the expanding schedule of practice it selects.

Basic Procedures

- Japanese-English vocabulary delivered on a computer
- 2 Trial Types
 - Study – Present cue and response
 - Recall or Study
 - Results in correct response or Study trial if failure
 - Model counts this as 1 practice either way
 - Success or failure provides information used for tracking learning
 - Trials have variable latency (faster and less chance of study if well learned)
- Experiment introduces items with Study trials
 - Subsequent trials are Recall or Study trials
- Continuous paired-associates
 - Trials mixed together according to:
 - Conditions set up by investigator
 - Decisions of optimization algorithm

Experiment

- 60 subjects
- 100 word pairs
- Session 1 — 1 hour of training
- Session 2 — 3 Recall or Study trials for each item
- 2 between-subjects factors on Session 1
 - Keyword Mnemonics Training or Free Strategy Instructions
 - Optimization or Widest Possible Spacing

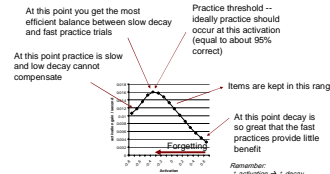
Optimal Scheduling of Practice

- Optimal scheduling maximizes the efficiency of the schedule. The amount of learning for a given cost of learning is maximized.
 - Amount of learning for a trial can be measured as a change in activation after a retention interval. (Learning can only be measured after a retention interval.)
 - Cost of learning for a trial is just the time for practice now.
 - We can compute a measure of efficiency of practice and call it activation gain per second (it is a sort of prescient learning rate).

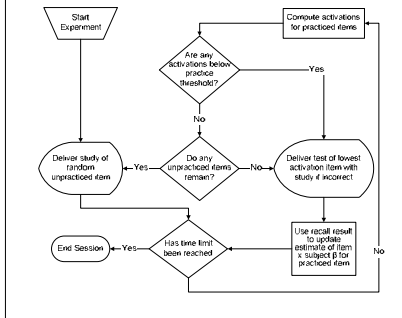
$$\text{activation gain per second} = \frac{\text{activation gain after retention interval for practice now}}{\text{time cost for practice now}}$$

- Activation gain per second depends only on activation (given fixed parameters and retention interval).
- And there is an activation at which it is maximal.
- All practice should occur at this optimally efficient activation level.

- Activations are computed for all items prior to each trial in the Optimization condition.
 - Forgetting means the items are constantly losing activation.
 - When an item falls to the practice threshold (the activation value at which practice is optimal) it is practiced.

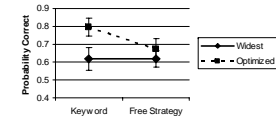


Implementation Flowchart

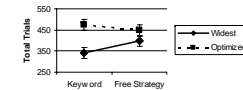


Results

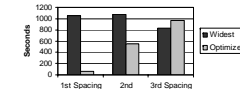
- Optimized scheduling was significantly superior to Widest Possible Spacing for Session 2 average performance, $F(1, 56) = 4.51, p < 0.05$.
- Main effect of Keyword Mnemonics Training and the interaction were also both significant. (However, not until a covariate was factored out to reduce variance.)



- These effects can be largely explained by the number of Session 1 trials completed in each condition.



- An interesting consequence of applying the algorithm is an expanding schedule on Session 1.
 - This is because as practices accumulate activation becomes more stable and it takes longer for an item to be forgotten to a 95% chance of recall.



Conclusion

Compared to cycling the items with the widest possible spacing, the optimization algorithm resulted in better scheduling decisions as measured by final performance.

The algorithm achieves its success because it results in more practices in the optimized conditions. These additional practices more than make up for the fact that wider spacing results in better learning per trial.

References

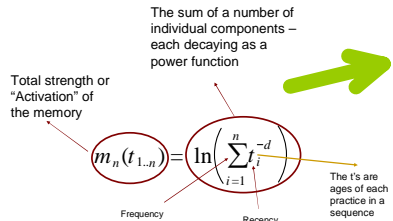
Anderson, J. R. & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
Pavlik Jr., P. I., & Anderson, J. R. (2003). An ACT-R model of the spacing effect. In F. Dieke, D. Dörner & H. Schaub (Eds.), *Proceedings of the Fifth International Conference of Cognitive Modeling* (pp. 177-182). Germany: Universitäts-Verlag Bamberg.
Pavlik Jr., P. I. (in press). Lessons from paired-associate learning. *Order Effects in Learning*. Oxford University Press.

Acknowledgments

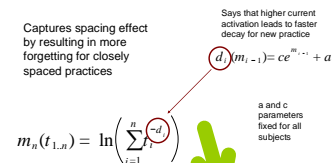
Preparation of this paper was supported by grant R01 MH68243 from the National Institute of Mental Health.

ACT-R Model Applied (Anderson & Lebiere, 1998; Pavlik & Anderson, 2003; Pavlik, in press)

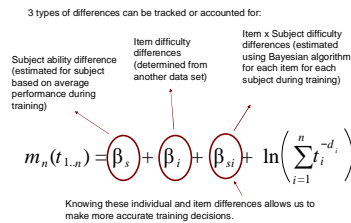
Base Level Activation Equation



Version with Decay Equation



Version used to track Individual and Item Differences



Activation over time



Output Functions

