

An ACT-R Model of Memory Applied to Finding the Optimal Schedule of Practice

Philip I. Pavlik (ppavlik@andrew.cmu.edu)

Department of Psychology, Carnegie Mellon University
Pittsburgh, PA 15213 USA

John R. Anderson (ja+@cmu.edu)

Department of Psychology, Carnegie Mellon University
Pittsburgh, PA 15213 USA

When a person confronts the task of memorizing a collection of facts some questions have to be answered about how to optimize their learning. Perhaps the most important of these questions is how the practices for each item should be scheduled. The most basic suggestion from psychology is that the practices should be spaced as widely as possible. Since Bahrick (1979) it has been clear that wide spacing intervals result in better performance after a long delay. However, while wide spacing is likely to help, it may also be important to consider item differences when making practice scheduling decisions.

Memory items (such as the Japanese-English paired-associates we have used in this research) are not equally difficult for each subject to learn. In postexperimental interviews subjects often mention that some items were easier to remember because they could find some mediator to connect the cue to the response. Further, the data shows that some items are on average easier for all subjects. Perhaps the cue (the Japanese word) for these items is fairly close to some common English word that can be used to mediate between the cue and response.

It is useful to consider how we might optimize learning by attending to this variability in item difficulty. For instance, if we knew an item was easier we could schedule it less often, while if it was more difficult we could schedule it more often. This may have two benefits. First, if we did not use this procedure and instead gave the items equal practice we might expect that some items would be overlearned (the easier items) and some items would be learned poorly (the harder items). This uneven learning might be undesirable. Second, by saving practices on easy items and then distributing them to hard items it might be possible to achieve gains in performance.

To attend to this variability requires a model of memory that does two things. First, it needs to describe accurately how practice in general leads to recall performance by characterizing memory effects due to the frequency, recency, and spacing of practices. Second, to optimize performance for individual items, the model needs to be able to characterize both initial average difficulty for each item (estimated by fitting data from prior experiments, from now on referred to as item difficulty) and the difficulty particular subjects have with each item (which will need to be estimated based on performance with the item over the

course of each subject's learning, from now on referred to as subject/item difficulty).

Model

The model we used, an extension of the Adaptive Character of Thought – Rational (ACT-R -- Anderson & Lebiere, 1998) theory, captures the three major effects in memory. The original ACT-R model captures the recency and frequency effects. Our extension of ACT-R (Pavlik and Anderson, 2003) captures the spacing effect. Importantly, it also models the spacing by practice interaction (that more spaced practice causes a larger spacing effect) and the spacing by retention interval interaction (that a longer retention interval causes a larger spacing effect).

In the model, memory strength for an item is represented by a value called activation which is calculated from the frequency, recency and spacing of practices. To capture the item difficulty for any item (both the item and subject/item components) we added a parameter to the model, β , which is added to activation for each item to represent whether the item is easy (positive β) or difficult (negative β). Overall β for any item is the sum of the item and subject/item components. A sigmoidal output function is used to calculate probability of recall as a function of activation.

Optimization

Our optimization routine supposes that training the lowest activation item leads to maximum learning (in terms of activation at a delay). This assumption follows from the model which describes how the long-run activation contribution of each practice depends on the activation for the item at the time the practice occurs.

The optimization procedure for a set of paired-associate items begins by presenting both the cue and response for each pair in a sequence of study trials. Following this, each trial involves the following steps:

1. Calculate the activation for all items.
2. Practice the item with the most to gain at recall (lowest activation item).
3. Update estimate of subject/item β for the practiced item.
4. If all trials completed end, else go to step 1.

The initial introduction of the items is in order from the items with lowest β s to the items with the highest β s. After each item has been introduced it has a history and the

routine can calculate its activation, which includes the current β estimate for the item (the sum of item and subject/item components). After computing the activations for all items, the lowest activation item is selected and practiced. Each practice is a test of the item, which is followed by the correct response if the subject supplies the wrong answer. Defining a practice in this way allows us to both give practice and assess performance with each trial.

The routine uses the performance from each trial to update the estimate of subject/item β for the item. The routine starts out by assuming a prior distribution $f(x)$ for subject/item β s determined in the prior experiment (Pavlik and Anderson, 2003). Each time the item is tested the mean and variance of this distribution are updated. If S is the history of success and failure for an item the posterior distribution is calculated according to the Bayesian formula:

$$f(x|S) = \frac{f(x)P(S|x)}{\int f(x)P(S|x)dx} \quad (1)$$

where $P(S|x)$ is the probability of S given a value x .

This procedure allows us to estimate the current subject/item β estimate for an item as a function of the history of performance with the item. Essentially, what this involves is successive shifting and narrowing of our subject/item estimate of β as we gather data for an item.

Experiment

As a test of this optimization algorithm, we compared it with 2 control conditions in a within-subjects design. The first control condition scheduled practice using the widest possible spacing, while in the second control condition the schedule was yoked to the schedule produced for the previous subject's optimized words (the schedules did not correspond to the same words for the previous subject.)

The experiment also had 2 between-subjects conditions. In addition to the Japanese-English pairs from the prior experiment, we also used a set of Spanish-English pairs. While the Japanese stimuli had been designed to prevent easy associations, the Spanish language set contained more variability. While some words were as difficult as Japanese, roughly half of the Spanish set was rather easy with close cognates in English.

Subjects learned 60 pairs in the Japanese condition and 90 pairs in the Spanish condition. The word pairs for each subject were randomized into the three conditions so one-third of the pairs was designated in each condition. Since we had estimates of item β s for Japanese pairs, the optimization used these as estimates of the initial item β s. We had no initial item β estimates for the Spanish pairs so assumed 0 values for all these items.

The experiment took place in two sessions separated by 2 days. On session 1, subjects went through 360 trials in which the 3 within-subject conditions were interleaved. On the second session all items were tested twice to determine the effects of conditions.

Because there was little difference between the between-subjects conditions they were aggregated for analysis. As

expected, both the wide spacing and optimized conditions (M s = 0.798 and 0.818) performed significantly better than the yoked condition (M = 0.726). The difference between the optimized and wide conditions was not significant.

To explore the results of the optimization, for all subjects within each within-subjects condition, we calculated the correlation of first session percent correct for each word with second session percent correct for each word. This gave three correlations for each subject, one for each condition. These correlations provided a measure of how likely a word that was responded to incorrectly on session 1 would be responded to incorrectly on session 2. With even or yoked spacing one might expect the correlation would be substantially positive because hard items would still be on average harder on session 2. In the optimized condition however, to the extent that hard items were given more practices and easy items fewer, one should expect that session 1 and 2 performance should be uncorrelated.

Since 10 subjects got 100% correct on session 2 for at least one of the conditions it was impossible to calculate one or more correlations. Statistical tests were performed on the 30 subjects with complete data. After showing the overall difference was significant using a within subjects ANOVA, $F(2, 58) = 52.6, p < .001$, we ran several t-tests to compare mean correlations (M s 0.54, 0.13 and 0.64 respectively for even, optimized and yoked conditions). Both even spacing and the yoked condition correlations were significantly greater than optimized, t s $> 6, p$ s < 0.001 .

Conclusion

The results of the experiment showed that the optimization routine behaved as expected. While the improvement was not significant compared with even spacing, the correlations showed that the optimization as it currently stands has the desirable property of automatically identifying difficult items and giving them more practice at the expense of easier items. This property might be especially useful in longer term training protocols during which extensive overtraining of well learned items would result in considerable inefficiency.

Acknowledgments

Preparation of this paper was supported by grant N00014-96-01-1491 from the Office of Naval Research.

References

- Anderson, J. R. & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ, US: Lawrence Erlbaum Associates. Publishers.
- Barrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General*, 108, 296-308.
- Pavlik Jr., P. I., & Anderson, J. R. (2003). An ACT-R model of the spacing effect. In F. Detje, D. Dörner & H. Schaub (Eds.), *Proceedings of the Fifth International Conference of Cognitive Modeling* (pp. 177-182). Germany: Universitäts-Verlag Bamberg.