Using a model for learning and memory to simulate learner response in spaced practice

Mark A. Riedesel[1*], Neil Zimmerman[1], Ryan Baker[2], Tom Titchener[1], and James Cooper[1]

[1] McGraw-Hill Education, 281 Summer Street, Boston, MA 02210
{mark.riedesel,neil.zimmerman,tom.titchener,james.cooper}@mheducation.com
[2] University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104
rybaker@upenn.edu

Abstract. McGraw-Hill Education's new adaptive flashcard application, StudyWise, implements spaced practice to help learners memorize collections of basic facts. For classroom use, subject matter experts needed a scheduling algorithm that could provide effective practice schedules to learn a pre-set number of facts over a specific interval of days. To test the pedagogical effectiveness of such schedules, we used the ACT-R model of memorization to simulate learner responses. Each schedule has one 30 minute study session per day, with overall study intervals that ranged from one day for sets of less than 30 items to three weeks for sets of two hundred or more items. In each case, we succeeded in tuning our algorithm to give a high probability the simulated learner answered each item correctly by the end of the schedule. This use of artificial intelligence allowed us to optimize the algorithm before engaging large numbers of real users. As real user data becomes available for this application, the simulated user model can be further tested and refined.

Keywords: Spaced Practice, LearnSmart, StudyWise, Adaptive Flashcards, Mobile Learning, iOS, Android

1 Introduction

For many subject areas, memorizing basic facts is an important first step in learning and mastering content. Examples include foreign language vocabulary, medical terms, and anatomy and physiology. Research over more than a century has shown that an effective way to memorize basic facts is through the use of spaced practice [2].

Applications designed for long-term memorization are often conceptually based on models for human memory that grew out of early work on how memories decay with time but can be reinforced by repetition spaced in time [3]. These include most existing commercial and open source adaptive flashcard applications. These are typically designed for learning large amounts of material (thousands of facts) over an extended period of time (weeks, months, or even years). Such applications include SuperMemo, Anki, Duolingo, Brainscape, and

Memorang [http://www.supermemo.com, http://ankisrs.net, http://www.duolingo.com, http://www.brainscape.com, http://www.memorangapp.com].

Significantly refined models for human memory developed over the past thirty years have been used to construct optimized schedules for spaced practice [6, 7, 5]. Research studies on this topic typically have two or three practice sessions separated by a day with one recall test at some specified time later, usually about a week [2, 7]. The schedules created are designed to optimize the time between practice sessions as a function of the time between the last practice session and the recall test. These models have been tried in the classroom but are not yet in widespread use [4].

Based on the research done in the past fifteen years or so, it has also been found that a pattern of spaced practice designed to fit within the time constraints of an academic class can significantly enhance learning, even if the schedule is not optimally derived from a cognitive model [1]. In this case, the challenge is to find an algorithm to produce a schedule for spaced practice that will result in effective retention of the material by the learners while still fitting within the time constraints of the course schedule. Artificial Intelligence in the form of cognitive models can be used to design and test such schedules even if the equations that describe the models are not directly used to construct the schedules themselves.

2 StudyWise

The Higher Education division of McGraw-Hill Education (MHE) wanted to use MHE's new adaptive flashcard application, StudyWise, to facilitate memorization of existing educational content being used in college courses. To do this, StudyWise presents questions, known as probes, that come from MHE's existing LearnSmart [http://www.mheducation.com/highered/platforms/learnsmart.html] database of probes.

In LearnSmart, each probe is associated with a Learning Objective (LO). The LOs are organized by Topic, which in turn are related to a LearnSmart title's subject. There are currently about 1500 LearnSmart titles on a wide range of subjects. In a course that uses LearnSmart, the instructor creates a LearnSmart assignment for an instructor specified set of LOs. Students see only probes associated with the LOs for that assignment, which they do on-line.

StudyWise presents all of the LOs associated with a particular Topic and uses its spaced practice algorithm to present probes associated with those LOs to the learner. The algorithm is designed to allow the learner to master each LO by repeated practice. These LOs are associated with the individual topics in five existing MHE LearnSmart titles. These titles are Introductory Spanish, Anatomy and Physiology, Medical Assisting, Human Resources, and Medical Terminology. The app is an entirely mobile one and has IOS and Android versions.

2.1 Study Schedules for Sets of Learning Objectives

The number of LOs for each subject bundle (deck) and the desired time to cover this material was specified by the Subject Matter Experts (SMEs) for each area.

They intend that learners use the app for a half an hour a day, four to five days a week. The desired schedules for covering the material were the following:

Table 1. Study Schedule by Deck Size

Number of LOs in the Deck	Total Hours of Study	Number of 30 Minute Sessions (Total Study Interval)
15-50	1-2 Hours	2-3 ($<$ 1 week)
51-100	2-3 Hours	4-6 ($\sim$ 1week)
100+	3+ Hours	7+ ($\sim$ 2 weeks)

2.2 Time per Probe from LearnSmart Data

For two of the LearnSmart titles from which the LOs were taken, we have data indicating how long students took to respond to each probe. This comes from college classes which have used these two LearnSmart titles. For these, we had data for 3,000,000 answers to about 1,500 different probes. This data indicated that an median response time of 15 seconds per probe was reasonable, giving a possible 120 probes in a 30-minute session.

2.3 StudyWise Spacing Algorithm

The challenge, then, was to find a spacing algorithm that could meet the following criteria: (1) distribute the appearance times for each LO within the targeted overall practice period for a given deck size, (2) have enough appearances for each LO that the learner will know it by the end of the practice interval, assuming no previous knowledge of the LO, and (3) not repeat a given LO more times than needed to learn it (i.e. don′t waste students′ time).

The starting point for the spacing algorithm used by StudyWise came from ALEKS QuickTables [`http://www.aleks.com/k12/quicktables`], which is used to teach elementary school children arithmetic tables for the numbers from 0 to 12. Hence it is optimized for a deck of 13x13 = 169 items or less. For StudyWise, we extended the QuickTables algorithm to allow the use of information on learner confidence for each probe and for the difficulty of each probe.

We needed to vary the parameters of the StudyWise algorithm to see if it was flexible enough to produce workable schedules for the range of deck sizes and the time constraints desired by the SMEs. To find viable schedules, we needed to simulate a learner who started with no initial knowledge of each LO but whose memory would improve with each repeat appearance. This would allow us to test a wide range of algorithm parameters and find the most effective algorithm parameters which might meet the SME′s criteria.

3 Simulating Memorization Using the ACT-R Model

To model a learner memorizing new material, we chose the ACT-R-based memory model of Pavlik and Anderson [3,4]. This model can calculate the probability that a simulated learner would remember a given LO at each appearance of it in the learning sequence. The initial activation level (strength in memory) for each item is set at zero, so the simulation assumes no prior knowledge. With each appearance, the activation level increases but then immediately starts to decay in accord with the model. The memory strength decreases both because of interference between LOs within a learning session and the decay of memory with time between sessions. Parameters used for the ACT-R model were those given in Pavlik and Anderson [6].

With these assumptions the user always gets the answer wrong on an LO's first appearance, since they have no initial memory of it. On subsequent appearances, the probability of remembering the LO is calculated using the Pavlik and Anderson model. A random number weighted by that probability is then calculated to decide if the user actually remembered the LO at that time.

We varied the spacing algorithm′s parameters to tailor the spacing and frequency of the appearance of each LO to try and fit the SME′s desired learning windows. Random numbers are used in the algorithm's method of selecting an LO for presentation to the learner and within the simulation to decide if the learner has actually remembered an LO at a given appearance. We therefore did 100 runs for each LO to find the range of variation in the pattern of appearances for a set of LOs from a given deck size. If the simulated user was calculated to have not remembered an item, the StudyWise algorithm repeats that item at some time later until it is answered correctly, on a repeat schedule that is part of the algorithm. For an item to be considered finished it must be answered correctly on its last appearance.

The question, then, was this: Could we find, for each deck size, a parameterization for the algorithm for which the simulated user, over 100 trials, successfully completes the entire set of LOs within at least close to the specified practice interval? It was not obvious at the outset that this would be possible.

Happily, though, we were able to find algorithm parameters that fit the SME′s specifications for time of study vs. deck size for all three cases in Table 1. This indicates that the SMEs intuition for how much a learner can memorize within a certain time interval agrees very well with how ACT-R models human learning and memorization. Simulation results for several deck sizes are given below.

3.1 Simulations for Decks of Varying Size

For a deck with 30 LOs, the tests indicated that all the LOs could be learned in less than two 30 minute sessions. Figure 1 shows the appearance time for an LO on the x-axis and the LO number on the y-axis. If an item was answered incorrectly, as determined by the simulation, the point is plotted as a red dot. If answered correctly, the point is plotted in blue. All thirty LOs are blue at the end of practice and were completed within two sessions, as specified in Table 1.

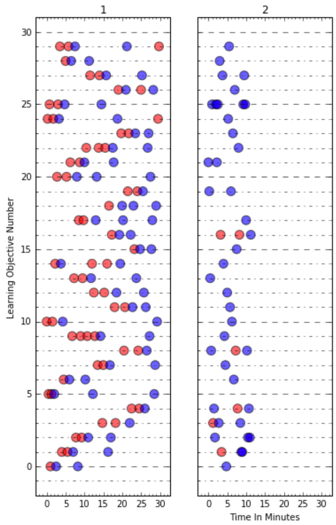

Fig. 1. LO number within a deck versus time of appearance for a deck of 30 items. A red dot indicates that the answer was incorrect, as calculated by the model, and a blue dots is for a correct answer. Each numbered block is a separate 30 minute session. The start times of the sessions are separated by 24 hours.

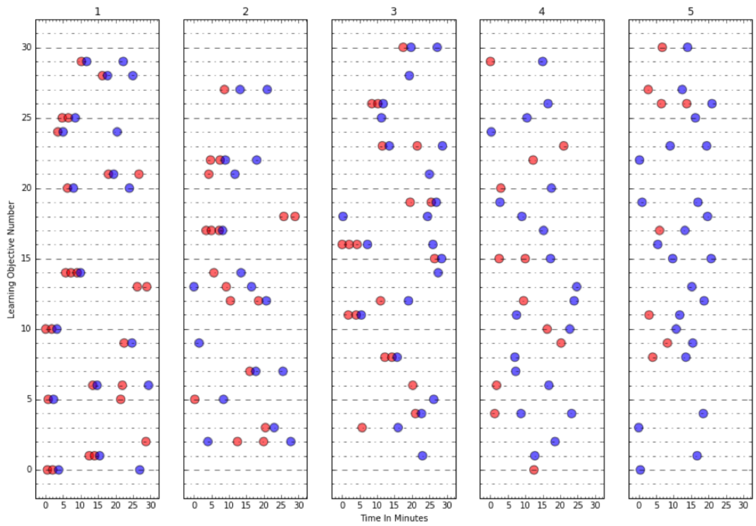

Fig. 2. Plots of the first 30 LOs from a 100 LO deck. Five 30 minute sessions were needed to complete 100 LOs. Red dots indicate incorrect answers, as computed by the model, and blue dots are for correct answers. The start times for the sessions are separated by 24 hours.

For 100 LOs, five 30 minute sessions were needed [Fig. 2]. All 100 of the LOs were answered correctly by the simulated user by the end of the last session. For 200 LOs eleven 30 minute sessions were needed, just slightly in excess of the SME's target of two weeks.

4 Future Work: Comparison with User Data

The application is instrumented to anonymously record data about a user′s session that is stored locally and then retrieved when the user has an internet connection. As this application is more widely deployed, we will be able to get direct feedback from the users and also to evaluate its effectiveness at helping users memorize the material and improve their performance in a class.

In sum, we have used artificial intelligence to develop an algorithm for MHE's adaptive flashcard applicaiton, StudyWise, which implements spaced practice to help learners memorize content connected to Learning Objectives in several existing MHE LearnSmart titles. Going forward, we can expand StudyWise for use with a wider range of subject areas and a more diverse set of content sources.

Acknowledgments. The target practice schedules were provided by Katie Ward and the MHE Higher Ed team of SMEs. The original QuickTables algorithm was developed by Jean-Claude Falmagne and Eric Cosyn.

References

1. Brown, P.C., Roedinger, H.L., McDaniel, M.A.: make it stick: The Science of Successful Learning. Belknap Press: An Imprint of Harvard University Press, Cambridge, MA (2014)
2. Cepeda, N.J., Vul, E., Rohrer, D., Wixted, J.T., Pashler, H.: Spacing effects in learning, a temporal ridgeline of optimal retention. Psychological Science 19(11), 1095–1102 (2008)
3. Ebbinghuas, H.: Memory, a contribution to experimental psychology. Dover, New York (1885)
4. Lindsey, R.V.: Probabilistic Models of Student Learning andForgetting. Ph.D. thesis, University of Colorado at Boulder (2014)
5. Mozer, M.C., Lindsey, R.V.: Predicting and improving memory retention: Psychological theory matters in the big data era. In: Jones, M. (ed.) Big Data in Cognitive Science. Taylor & Francis (2016)
6. Pavlik, P.I., Anderson, J.R.: Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. Cognitive Science 29, 559–586 (2005)
7. Pavlik, P.I., Anderson, J.R.: Using a model to compute the optimal schedule of practice. Journal of Experimental Psychology: Applied 14(2), 101–117 (2008)