A Queueing Network Model for Spaced Repetition

Siddharth Reddy
Dept. of Computer Science,
Cornell University
Ithaca, NY 14850, USA
sgr45@cornell.edu

Igor Labutov
Dept. of Electrical and
Computer Engineering, Cornell
University
Ithaca, NY 14850, USA
iil4@cornell.edu

Siddhartha Banerjee
School of Operations Research
and Information Engineering,
Cornell University
Ithaca, NY 14850, USA
sbanerjee@cornell.edu

L@S 2016, April 25–26, 2016, Edinburgh, UK.
ACM 978-1-4503-3726-7/16/04.
http://dx.doi.org/10.1145/2876034.2893436

Abstract

Flashcards are a popular study tool for exploiting the spacing effect – the phenomenon in which periodic, spaced review of educational content improves long-term retention. The Leitner system is a simple heuristic algorithm for scheduling reviews such that forgotten items are reviewed more frequently than recalled items. We propose a formalization of the Leitner system as a queueing network model, and formulate optimal review scheduling as a throughput-maximization problem. Through simulations and theoretical analysis, we find that the Leitner Queue Network (LQN) model has desirable properties and gives insight into general principles for spaced repetition.

Author Keywords

Spaced Repetition; Flashcard Scheduling

Introduction

The ability to retain a large number of new ideas in memory is an essential component of human learning. In recent times, there has been a growing body of work that attempts to 'engineer' this process – creating tools that enhance the learning process by building on the scientific understanding of human memory. Flashcards are one such tool that use the idea of *spaced repetition* to overcome the human 'forgetting curve'. Though they have been around for a while in the physical form, a new generation of spaced repetition software such as SuperMemo [9], Anki [4], and Mnemosyne [1] allow a much greater degree of control and monitoring of the process. As these software applications grow popular, there is a need for formal models for reasoning about and optimizing their operations. In this work, we use ideas from queueing theory to develop such a formal model for one of the simplest and most popular spaced repetition systems: the Leitner system.

Related Work

The exponential forgetting curve, which was first studied by Ebbinghaus in 1885 [3], models the probability of recalling an item as a function of the time elapsed since previous review and memory 'strength'. The exact nature of how strength evolves as a function of the number of reviews, length of review intervals, etc. is not clear, though a general *spacing effect* in which spaced reviews lead to greater strength than massed reviews (cramming) has been observed [2]. Recent studies have proposed more sophisticated probabilistic models of learning and forgetting [8, 6]. In our queueing model, we assume the exponential forgetting curve and a simple model of memory strength.

Novikoff et al. have proposed a theoretical framework for spaced repetition [7] that assumes strict spacing constraints and considers items to be identical. They propose deterministic algorithms for satisfying different spacing constraints, and examine the rate at which new items can be presented to different types of students using these algorithms. We improve upon their ideas by posing the throughput-maximization problem as an optimization problem that incorporates the user's review frequency budget and non-identical item difficulties.

Memory Model

The probability of a student recalling an item is as follows.

$$\mathbb{P}[\text{recall}] = \exp\left(-\theta \cdot d/s\right)$$

where θ is the item difficulty, d is time elapsed since previous review, and s is memory strength. In the queueing model, we will assume that memory strength s is equal to the position of the deck that the item is currently in. We have performed preliminary experiments on large-scale log data from the popular Mnemosyne spaced repetition software that validate this probabilistic model of memory.

Queueing Network Model

Leitner System

The Leitner system is a heuristic for prioritizing items for review. After the user sees an item for the first time, it enters the system at deck 1. Each deck is a first-in-first-out queue, and when the user requests an item to review, the system randomly chooses a deck i, and shows the user the item at the top of deck i. If the user forgets the item, it is added to the bottom of deck $i-1$. If the user recalls the item, it is added to the bottom of deck $i+1$. The key idea is that when the system randomly chooses a deck, it places more weight on lower decks than higher decks, so the user spends more time working on new and difficult items, and less time on items that are almost mastered. A user might work on items from deck 1 every day, deck 2 every other day, deck 3 every week, deck 4 every month, etc. One of our main contributions is a principled method for selecting deck review frequencies to maximize the rate at which items are mastered while respecting the user's review budget.

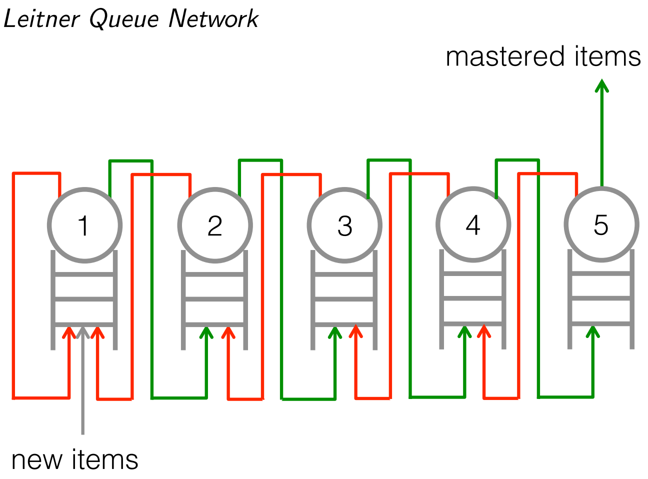

Figure 1: A diagram depicting the routing of items between queues in the network model (where the number of queues is $n = 5$). Green arrows indicate transitions that occur when an item is recalled, and red arrows indicate transitions for forgotten items.

Consider a network of n inter-connected M/M/1 queues [5], as in Figure 1. New items arrive into deck 1 according to a Poisson process with rate λ_{ext}. The routing probability matrix is P, where $P_{ij} = \mathbb{P}[\text{recall} \mid s = i, \cdot]$ when $i < n \wedge j = i + 1$, $P_{ij} = 1 - \mathbb{P}[\text{recall} \mid s = i, \cdot]$ when $(i > 1 \wedge j = i - 1) \vee i = j = 1$, and $P_{ij} = 0$ otherwise.

Items exit the system from deck n with probability $\mathbb{P}[\text{recall} \mid s = n, \cdot]$. The service rate for deck i is indicated by μ_i, and the user's work rate budget (e.g., the maximum number of items the user can review per day) is given by U. This network of queues is a *Jackson network*, and can be treated as a continuous-time Markov chain. Thus, we can characterize the steady state of the system using flow-balance equations and queue length stability conditions. We are interested in finding μ_i that maximize the steady-state throughput of the system such that the budget constraint $\sum_{i=1}^{n} \mu_i \leq U$ is satisfied. Formally, we must solve the following *static planning problem*.

$$\begin{array}{lll} \text{maximize} & \lambda_{ext} & \\ \text{subject to} & \displaystyle\sum_{i=1}^{n} \mu_i \leq U & \\ & \lambda_{ext} + P_{11}\lambda_1 + P_{21}\lambda_2 = \lambda_1 & \\ & P_{12}\lambda_1 + P_{32}\lambda_3 = \lambda_2 & \\ & \vdots & \\ & P_{n-1,n}\lambda_{n-1} + P_{nn}\lambda_n = \lambda_n & \\ & \mu_i, \lambda_i \geq 0 & \forall i \\ & \lambda_i < \mu_i & \forall i \end{array}$$

We can use an application of Jensen's inequality and the closed-form expected delay in a Jackson network to turn this into a tractable nonlinear optimization problem, and use a solver (e.g., IP-OPT) to find the optimal deck review rates μ_i^*. The algorithm for selecting the next item to present to the user is simple: sample deck i with probability $\frac{\mu_i}{\sum_{i=1}^{n} \mu_i}$, and select the item at the top of the sampled deck.

The vanilla model assumes a global item difficulty θ. We can easily extend the model to handle item-specific difficulties θ_i by creating parallel copies of the queueing system for different discretized difficulties, and enforcing joint budget constraint on the parallel systems.

Experiments

Exploring throughput-optimal review policies, we observe several intuitive results (e.g., the user should review lower decks more frequently than higher decks), as well as the following non-obvious results: (1) expected delay between consecutive reviews *increases* as an item moves up through the system – this is desirable behavior, since there is support in the experimental psychology literature for expanding intervals between repetitions [2]; (2) for small item difficulty, the user should spend a roughly uniform amount of time on each deck, but for large item difficulty, the user should spend more time on lower decks than higher decks; (3) maximum throughput λ^*_{ext} seems to be a convex function of review budget U, which implies *increasing* returns to user effort.

We simulated the queueing system and verified that a phase transition occurs when the arrival rate of new items exceeds the maximum throughput predicted by the Leitner Queue Network model: throughput decreases sharply for arrival rates greater than the predicted threshold.

Conclusion

We have proposed the first formalization of the Leitner system, which can be used to help spaced repetition software developers and users calibrate review schedules to achieve learning goals under a budget. Simple topological properties of our queueing network model lead to desirable properties in review schedules (e.g., expanding delays between reviews), and the model can be easily extended to handle realistic scenarios (e.g., non-identical item difficulties). Jupyter notebooks for running our experiments – as well as a full paper describing recent results – are available online at http://siddharth.io/leitnerq.

Acknowledgements

This work was funded in part by the Cornell Presidential Research Scholars Program, and a grant from the John Templeton Foundation provided through the Metaknowledge Network at the University of Chicago.

References

[1] The mnemosyne project. http://mnemosyne-proj.org, 2006.

[2] Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., and Rohrer, D. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological bulletin 132*, 3 (2006), 354.

[3] Ebbinghaus, H. *Memory: A contribution to experimental psychology*. No. 3. University Microfilms, 1913.

[4] Elmes, D. Anki. http://ankisrs.net, 2015.

[5] Kendall, D. G. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chain. *The Annals of Mathematical Statistics* (1953), 338–354.

[6] Lindsey, R. V., Shroyer, J. D., Pashler, H., and Mozer, M. C. Improving students' long-term knowledge retention through personalized review. *Psychological science 25*, 3 (2014), 639–647.

[7] Novikoff, T. P., Kleinberg, J. M., and Strogatz, S. H. Education of a model student. *Proceedings of the National Academy of Sciences 109*, 6 (2012), 1868–1873.

[8] Pashler, H., Cepeda, N., Lindsey, R. V., Vul, E., and Mozer, M. C. Predicting the optimal spacing of study: A multiscale context model of memory. In *Advances in neural information processing systems* (2009), 1321–1329.

[9] Wozniak, P., and Gorzelanczyk, E. J. Optimization of repetition spacing in the practice of learning. *Acta neurobiologiae experimentalis 54* (1994), 59–59.