A BAYESIAN MODEL FOR LEARNING USING FLASHCARDS

Venelin Valkov

University of Plovdiv venelinvulkov@gmail.com

***Abstract**: Memorising large amounts of unstructured information and vocabulary is required when studying foreign language, law, biology and medicine. Distributed over time review sessions benefit the long-term retention more than massed practice when studying such material. Flashcard learning using spaced repetition is one implementation of the distributed technique. This paper proposes a Bayesian bandit algorithm which tries to maximise the number of presented flashcards that the user is going to guess wrong in a study session. The suggested model is implemented in a mobile application.*

Keywords: spacing effect, multi-armed bandit, Thompson sampling, Bayesian learning

1. Introduction

Researchers have been trying to improve learning and retention since 19th century. One of the first to conduct experiments of human memory was Hermann Ebbinghaus [1]. His studies proposed methods for studying memory. He also discovered the spacing principle which suggests that having sleep periods between study sessions improve performance compared to contiguous sessions. Furthermore, the more often the learner encounters a piece of information the less often he needs to refresh it to keep it in memory.

Spaced repetition is a learning technique, which exploits the spacing principle. The learner is subjected to reviews of previously learned material in increasing intervals. One simple implementation of the spacing principle is Leitner's system [2]. It uses flashcards and has been incorporated in many spaced repetition software (SRS) programs. A flashcard is a card with information on either or both sides used during study sessions. The Leitner System can be presented as a box of flashcards with labeled compartments (e.g. 1 to 5). A flashcard is placed in the first compartment if it is still new. Those cards will be repeated every day. The second compartment contains flashcards that the learner knows relatively well. The cards change compartments when the learner knows them better. Every compartment has different repetition interval. In case of wrong answer the learner puts the flashcard in previous compartment.

One common problem in SRS is deciding when the user should study. A model based on Adaptive Character of Thought - Rational (ACT-R) was proposed to solve this problem (see [3] for background). This extended ACT-R model focuses on using set of equations that describe the strength of a memory chunk as a function of practice. Another problem is deciding which flashcard should be presented to a user at specific time for optimal learning performance. It is believed that repetition should enhance future performance, e.g. see [4]. This paper focuses on the first problem using a multi-armed bandit (MAB) modeling approach.

Since only one flashcard can be presented to the user at a time, the decision problem is sequential. Using only flashcards that the learner knows will not yield learning of all available information. Furthermore, the probability of getting a wrong answer for a flashcard that the user has not seen is unknown. Thus, a exploration/ exploitation trade-off exists. This paper uses the MAB problem setting to model the decision of picking which flashcard to show next. The payoff after each round is binary - the learner either knows the answer or he doesn't. This setting can be modeled using the Binomial Bandit (see [6]) which assumes that the payoffs of each arm are independent Bernoulli random variables with success probabilities $(\theta_1, \dots, \theta_k)$, where k is the number of arms. The goal of the model is to maximise the number of flashcards shown to which the user gives wrong answer for a given study session.

1.1. The multi-armed bandit problem

In a MAB problem, a player is presented with a sequence of slot machines. Each machine offers random reward from a distribution specific to that machine. In each round the player chooses from a set of alternatives ("arms") based on past history and receives the payoff associated with his decision. The goal is to maximize the total payoff of the chosen arms. This setting is often used to model situations where exploration/exploitation trade-off exists.

1.2. Bayesian strategy (Thompson sampling) for the MAB problem

Multiple strategies for finding approximate solution to the MAB problem exist [7]. Some of the most widely implemented are Upper Confidence Bounds (UCB) and ϵ-greedy. The Bayesian strategy (also known as Thompson sampling) is a probability matching strategy that is relatively easy to implement. The idea of the algorithm is to randomly draw each arm according to its probability of being optimal. Despite its simplicity this strategy achieves state-of-the-art results [8]. In the *K*-armed Bernoulli bandit setting the reward for the i-th arm is a Bernoulli distribution with mean μ_i^*. It is standard to model the mean reward of each arm using a Beta distribution since it is the conjugate distribution of the binomial distribution e.g. see [8] for background. The Beta distribution is defined on the interval [0, 1] and it is parametrized by two positive shape parameters, denoted by α and β that control the shape of the distribution. The parameters are chosen to reflect existing belief or information.

The total expected regret R is a popular performance measure for bandit algorithms, defined for round T as:

$$R_T = T\theta^* - \sum_{t=1}^{T} \theta_i\,(i)$$

where $\theta^* = max_{j=1,\ldots k}\theta_j$ is the expected reward for the best arm. An asymptotic lower bound was established for the algorithm (see [9])

$$R(T) \geq \log(T)[\sum_{i=1}^{K} \frac{p^* - p_i}{D(p_i||p^*)} + o(1)]$$

where p_i is the reward probability of the i-th arm, $p^* = maxp_i$ and D is the Kullback-Leibler divergence. A total regret of 0 means that the algorithm is achieving optimal performance, which is unlikely in practice. Good strategy's total regret should flatten as it learns the optimal arm to pull. The maximum payoff one can achieve at each round is by picking the arm with maximum probability of highest payoff.

2. Methods

Bernoulli bandit with Bayesian strategy was implemented. The success S of a single Bernoulli experiment was defined as showing a flashcard to which the learner gave wrong answer. The failure F was defined as showing a flashcard to which correct answer was given. The reward for a success outcome was set to 1 and that of a failure to 0. The proposed model tried to maximise the number of successes at each time step t based on previous $t - 1$ outcomes. The reward for flashcard i was modeled as a Beta-distributed random variable θ_i. $Beta(1,1)$, which is uniform on [0, 1], was chosen for prior distribution since no previous information about the learner's knowledge for any flashcard was present.

At time t having observed $C_i(t)$ correct and $N_i(t)$ incorrect answers the algortithm obtains the posterior distribution on θ_i

$$\mathrm{Beta}(1 + N_i(t), 1 + C_i(t))$$

for flashcard i. The decision of which card to show next was made by choosing the maximum θ_i drawed from each posterior distribution.

A sample run of the algorithm on a single flashcard is presented on *Figure 1*. The hidden probability for this example was set to $\theta = 0.7$ with a total of $T = 15$ rounds. The mean value of the posterior was 0.68 after completing all rounds.

The measure of total regret was used in order to quantify the performance of the algorithm. The maximum payoff one can achieve at each round T is showing flashcard i with maximum probability for success S.

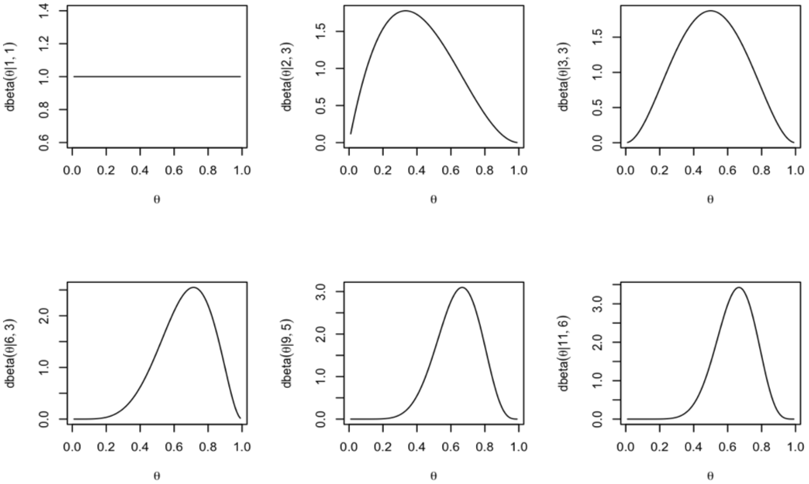

Figure 1. The upper-left figure shows the prior distribution of θ before running the algorithm. The bottom-right figure presents the posterior distribution for θ after showing the flashcard 15 times

The core algorithm was implemented as a server component in the programming language *Python* using the *SciPy* software library. The client was developed for the mobile operating system *iOS*. The communication between the components was performed via standard REST API.

3. Future work

Possible future developments include recommendations of specific study times, showing flashcards from other users, different study modes and automatic generation of concept maps.

Usage data from the mobile application will be collected from real users and made available for additional analysis. The practical performance of the model remains to be evaluated.

4. Conclusion

This paper proposed a model for creating flashcard software system using MAB model with Bayesian strategy. A way to evaluate it's performance empirically was provided. Two important simplifications has been made. Prior learner knowledge and study session times were not considered in the model. The model has been implemented in a mobile application for the *iOS* operating system.

Acknowledgements

This study is partially supported by project "SP15-FMIIT-015/11.05.2015". I would like to thank Veska Noncheva for her constructive feedback and general help with the research for this work. I would also like to thank to Polina Zhelyazkova for her proofreading work and continuous support. Finally, I would like to thank the reviewers of this work.

References

1. Ebbinghaus, H. (1885). Über das gedächtnis: untersuchungen zur experimentellen psychologie. Duncker & Humblot.
2. Leitner, S. (1974). So lernt man lernen. Herder.
3. Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. Journal of Experimental Psychology: Applied, 14(2), 101.
4. Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. Trends in cognitive sciences, 10(1), 14-23.
5. Agrawal, S., & Goyal, N. (2011). Analysis of Thompson sampling for the multi-armed bandit problem. arXiv preprint arXiv:1111.1797.
6. Scott, S. L. (2010). A modern Bayesian look at the multi- armed bandit. Applied Stochastic Models in Business and Industry, 26(6), 639-658.
7. Kuleshov, V., & Precup, D. (2014). Algorithms for multi-armed bandit problems. arXiv preprint arXiv:1402.6028.
8. Chapelle, O., & Li, L. (2011). An empirical evaluation of thompson sampling. In Advances in neural information processing systems (pp. 2249-2257).
9. Lai, T. L., & Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. Advances in applied mathematics, 6(1), 4-22.

БЕЙСОВ МОДЕЛ ЗА ОБУЧЕНИЕ ЧРЕЗ ФЛАШКАРТИ

Венелин Вълков

Резюме: Запомнянето на голямо количество неструктурирана информация и лексикални значения на думи е задължително при изучаване на чужд език, право, биология и медицина. Разпределени във времето преглеждания подпомагат *дългосрочното запомняне повече от дълги учебни сесии при изучаване на подобен вид материал. Едно приложение на разпределеният подход е обучението чрез флашкарти с раздалечени повторения. Този доклад предлага Bayesian bandit алгоритъм, който се опитва да максимизира броя представени флашкарти на които потребителя ще даде грешен отговор в една учебна сесия. Предложеният модел е реализиран в мобилно приложение.*