Adaptive Quiz Generation Using Thompson Sampling

Fuhua Lin

Athabasca University, Canada
oscarl@athabascau.ca

Abstract. This paper proposes an approach to generating quizzes for a gamified formative assessment system, QuizMASter. To improve its performance and incorporate adaptivity to the system, the quiz sequence generation process is modeled as a Beta Bernoulli Bandit model and solving it with Thompson Sampling algorithm. Thompson sampling is selected because it is non-deterministic and can use prior knowledge. We test the effectiveness of the proposed multi-armed bandit algorithm in QuizMASter with an online course on Data Structure and Algorithms.

Keywords: Multi-armed bandits; Thompson sampling; Formative assessment; Gamification; Adaptive learning; Online learning; Education.

1 Introduction

The primary purpose of a formative assessment in online learning is to offer learners feedback through identifying areas that may need improvement. Currently, even though there are online self-quizzes in many online courses that provide immediate and ongoing feedback to students, those of 'one-size-fits-all', static, and manually made quizzes are boring to students. To make formative assessment in online courses engaging, we have been developing and testing a gamified formative assessment system, QuizMASter [1]. QuizMASter was designed similar to a TV game show, where a small group of contestants compete by answering questions presented by the game show host [1]. In QuizMASter, students naturally take the place of game contestants, however the game host has been replaced with a software agent. The issue with the current version of QuizMASter is its high complexity in quiz creation because of the time to create the student modeling, the necessity of calibration of the question bank. More seriously, it depends on knowledge of item difficulties to estimate students' proficiency while it assumes that the levels of difficulty remain unchanged over time.

To improve the performance of QuizMASter, in this research we aim to design an adaptive mechanism for automatically and dynamically generating engaging, and pedagogically helpful quizzes to be used in QuizMASter. To reach this goal, it is needed to design an algorithm that can accurately identify the lacking areas of knowledge of the student and explore the topic in detail helping further learning or remediation.

We model the quiz sequence generation process as a Beta Bernoulli Bandit model and solving it with Thompson Sampling algorithm [2]. Thompson sampling is selected

because it is non-deterministic and can use prior knowledge. We test the effectiveness of the proposed multi-armed bandit algorithm used in QuizMASter with an online course on Data Structure and Algorithms.

Section 2 will review the related work in adaptive formative assessment through using multi-armed bandits. Section 3 explains the method we proposed. The last section concludes the paper with future work.

2 Literature Review

MAB family of algorithms is named after a problem for a gambler who must decide which arm of a K-slot machine to pull to maximize his total reward in a series of trials [3]. These algorithms are capable of negotiating exploration-exploitation trade-offs. They have been applied in real-world applications solving optimization problems such as educational experimental design [4] and website optimization [5]. Recently, there are emerging applications of MAB algorithms for optimal learning material selection [6-7]. Recently, Melesko and Novickij (2019) [8] propose and test an alternative adaptive testing method based on one of the Multi-armed Bandit (MAB) algorithms, Upper-Confidence Bound [9] for formative assessment for computer networking and cloud computing technologies. UCB algorithm is selected as it is one of the simplest algorithms that offers sub-linear regret. The algorithm suggests choosing the action with the largest confidence bound. A topic with the smallest lower confidence bound is selected. Then the question number *n* chosen on learning objective *o* will be

$$o_n = arg\min_{t \in T}(\mu_o - C\sqrt{\frac{logn}{N_o}}) \quad (1)$$

Where *C* is a constant that can be chosen to regulate the impact the second exploration has on the choice of the learning objective, and N_O is the number of questions for which the learning objective *o* has been asked so far. As the number of questions on the learning objective increases, so the uncertainty and the exploration term of the formula decrease. They obtained some initial positive results [8].

However, there are three drawbacks of UCB algorithm. First, it is a deterministic algorithm. If the input to the algorithm is the same, the output will be the same. This feature is not desirable when the student plays the game repeatedly. Second, it cannot use prior knowledge about the mastery level of the student while prior knowledge may be useful to identifying the lacking areas of knowledge of the student. Third, according to [8], the number of questions needed to achieve a certain level of accuracy is a function of exploration constant C in (1), which is unknown to us.

Thompson sampling (also called posterior sampling [2] strategy was first proposed in 1933 by Thompson [10] but it attract little attention in literature on MAB until recently when researchers started to realize its effectiveness in simulation and real-world applications [11-13]. Its main idea is to randomly select an arm according to the probability that it is optimal.

3 The Proposed Method

3.1 The Quiz Model in QuizMASter

We represent the domain model of a subject as $\Delta = \{\delta_1, \delta_2, \ldots, \delta_n\}$, δ_i is called knowledge unit (KU). Each *KU* has a list of learning objectives. The learning objectives of δ_i are denoted as $LO(i) = \{lo(i,1), lo(i,2), \ldots, lo(i,j), \ldots lo(i,n_i)\}$. ($i = 1, 2, \ldots, K$). Here n_i is the number of learning objectives for knowledge unit δ_i , and $lo(i,j)$ is j^{th} learning objective in δ_i. For $lo(i,j)$, we design a set of assessment questions. For each assessment question, it corresponds to one or more learning objectives and one or more *KU*s.

A quiz in QuizMASter consists of a set of questions which could be multiple-choice questions (MCQs) or True/False questions, denoted as $Quiz = \{q(1), q(2), \ldots, q(i), \ldots q(m)\}$, in which $q(m)$ is the size of a quiz, which is to be determined optimally. Each MCQ has several options, only one of which is correct. Each MCQ is tagged by the course experts with a set of indexes according to the assessment model, including corresponding KUs and learning objectives and feedback. The student model is to estimate the probability that a given student with a history will answer to a test question associated with a learning objective correctly. It is represented as a time-series matrix where rows represent the learning objectives, columns represent discrete times, and the value is the probability that the student can answer the questions of the learning objective correctly. We record all the answers (correct or wrong) the student answered for each question.

3.2 Modelling the Quiz Generation Process with Thompson Sampling

The accuracy of identifying the weakest learning objectives of the student is important for two reasons. One is for ensuring that the game is challenging and engaging and another is for providing accurate and helpful feedback to the student. Considering the limited time of the learner and limited questions in the question bank, one of the objectives of the quiz generation process in QuizMASter is to maximize the accuracy of identifying the weakest learning objectives of the student in a quiz.

Thus, it is an optimization of sequential allocation problem. The host agent of the QuizMASter, representing the course instructor, explores the different learning objectives and engages in focused questioning, exploiting those learning objectives which are possibly in most need of further learning activities for remediation.

Bernoulli Bandit problem: In the Bernoulli bandit, there are *K* actions, and when played at time *t*, an action $k \in \{1, \ldots, K\}$ produces a reward r_t of one (1) with probability $\theta_k \in [0, 1]$ and a reward of zero (0) with probability 1-$\theta_k \in [0, 1]$. Each θ_k can be interpreted as an action's success probability or mean reward. The success probabilities $(\theta_1, \ldots, \theta_K)$ are unknown to the agent, but are fixed over time, and therefore can be learned by experimentation [2]. The objective is to maximize the cumulative number of successes over *T* periods, where *T* is relatively large compared to the number of arms *K*.

Bata-Bernoulli-Bandit: Let the agent begin with an independent prior belief over each θ_k. Take these priors to be beta-distributed with parameters $\alpha = (\alpha_1, \dots, \alpha_K)$ and $\beta = (\beta_1, \dots, \beta_K)$. For each action *k*, the prior probability density function of θ_k is

$$p(\theta_k) = \frac{\Gamma(\alpha_k + \beta_k)}{\Gamma(\alpha_k)\Gamma(\beta_k)} \theta_k^{\alpha_k - 1}(1 - \theta_k)^{\beta_k - 1}$$

Where Γ denotes the gamma function. Each action's posterior distribution is also beta with parameters that can be updated as follows:

$$(\alpha_k, \beta_k) \leftarrow \begin{cases} (\alpha_k, \beta_k), if\ x_t \neq k \\ (\alpha_k, \beta_k) + (r_t, 1 - r_t), if\ x_t = k \end{cases}$$

TS is specialized to the case of a Beta-Bernoulli bandit. The success probability estimate $\hat{\theta}_k$ is randomly sampled from the posterior distribution, which is a beta distribution with parameters α_k and β_k, rather than taken to be the expectation $\alpha_k/(\alpha_k + \beta_k)$ used in the greedy algorithm. $\hat{\theta}_k$ represents a statistically plausible success probability. We can see that the quiz generation problem in QuizMASter can be modelled as a Bernoulli Bandit problem.

As the goal is to explore the knowledge of the student, the algorithm should probe and explore the different topics and engage in focused questioning, exploiting those which are possibly in most need of further learning or remediation.

The host agent chooses from many questions from multiple learning objectives, which correspond to bandits in the multi-armed bandit model. Choosing a learning objective to explore is pulling an arm in the multi-armed bandit model.

The gambler is replaced by the game host. The choice of machines is replaced by a choice of an MCQ, and reward is replaced by the correctness of the answer by the student {0, 1}. The game host repeatedly chooses a question with a learning objective to explore until reaching the maximum number of a quiz. Our bandit model presents an opaque bandit problem where a unique answer, reward, is observed at each round.

Suppose the learning objectives to be assessed for the student are $LO = \{lo_1, lo_2, \dots, lo_K\}$. The reward in the QuizMASter to each question $x_r \in \{0, 1\}$ is binary valued. Each learning objective corresponds to an unknown probability distribution. There exists a vector $\mu \in [0, 1]^K$ such that for the r^{th} question of a quiz, the algorithm chose learning object lo_k the probability that $x_r = 1$ is $p(x_r = 1 | r; lo_k) = \mu_k$, k ∈ {1, 2, …, K}. That is, $\{\theta_k\}$ in the Bernoulli bandit is replaced by $\{\mu_k\}$. Take these priors to be beta-distributed with parameters $\alpha = (\alpha_1, \dots, \alpha_K)$ and $\beta = (\beta_1, \dots, \beta_K)$. For each action *k*, the prior probability density function of μ_k is

$$p(\mu_k) = \frac{\Gamma(\alpha_k + \beta_k)}{\Gamma(\alpha_k)\Gamma(\beta_k)} \mu_k^{\alpha_k - 1}(1 - \mu_k)^{\beta_k - 1}$$

To make the exploration of the weakest learning objective of the student's knowledge, the optimal policy is to choose a question on one learning objective for which μ_k attains its smallest value, i.e. $lo^* = argmin_{k \in K} \mu_k$.

3.3 The Proposed Algorithm

Considering the traits of UCB algorithm mentioned in Section 2, we adopt Thompson Sampling (TS) to solve the quiz generation online decision problem modelled as Bata-Bernoulli-Bandit problem. We cannot directly use standard model of TS, as it

ignores the process of prior specification and assumes a simple model in which the set of feasible actions is constant over time and there is no side information on decision context [2]. Through using a student model as prior knowledge, then TS can learn to select the quiz questions with the weakest learning objectives within the scope of a specific knowledge area to be assessed. Also, by imposing the pedagogical model and other game factors as time-varying constraints on the actions, i.e., the set of learning objectives. Table 1 shows the algorithm proposed to solve the quiz generation problem in QuizMASter.

Algorithm BernTS-QuizMASter(LO, α, β)

1: **for** t = 1, 2, …, **do**
2: # sample model
3: **for** k = 1, …, K **do**
4: Sample $\hat{\mu}_k \sim beta(\alpha_k, \beta_k)$
5: **end for**
6: #select and apply action:
7: $lo_t \leftarrow argmin_k \hat{\mu}_k$
8: Select a question with lo_t and observes x_t
9: #update distribution
10: $(\alpha_{x_t}, \beta_{x_t}) \leftarrow (\alpha_{x_t} + x_t, \beta_{x_t} + 1 - x_t)$
11: **end for**

We are developing a new version of the QuizMAster for experiments using the proposed framework. We organize the quiz game system for a course as a number of stages, each of which corresponds to a knowledge unit in the domain model. For instance, in Data Strucruee and Algorithms, the quiz game has 12 stages. In each quiz generation provess, the proposed algorithm can learn to select the quiz questions with the weakest learning objectives within the knowledge unit.

4 Conclusion and Future Work

We have presented a method for quiz generation for a quizmaster, a gamified and adaptive formative assessment system. With the method, we use a computer science online course on Data Structure and Algorithms as a testbed to test the feasibility and effectiveness of the proposed method. The research is at the initial stage and the system is still under development and to be tested by real online students from Athabasca University. We will use positive predictive value proposed in [8] as accuracy to measure the machine learning performance of the algorithm for benchmarking.

References

1. Dutchuk, M., Muhammadi, K. A., and Lin, F. **(**2009**),** QuizMASter - A Multi-Agent Game-Style Learning Activity, EduTainment 2009, Aug 2009, Banff, Canada, Learning by Doing, (eds.), M Chang, R. Kuo, Kinshuk, G-D Chen, M. Hirose, LNCS 5670, 263-272.

2. Russo, D. J., Roy, B. V., Kazerouni, A., Wen, Z., A Tutorial on Thompson Sampling, Foundations and Trends® in Machine LearningJuly 2018 https://doi.org/10.1561/2200000070
3. Vermorel J., Mohri M. (2005) Multi-armed Bandit Algorithms and Empirical Evaluation. In: Gama J., Camacho R., Brazdil P.B., Jorge A.M., Torgo L. (eds) Machine Learning: ECML 2005. ECML 2005. Lecture Notes in Computer Science, vol 3720. Springer, Berlin, Heidelberg
4. Rafferty, Anna N.; Williams, Joseph Jay; Ying, Huiji, (2019), Statistical Consequences of Using Multi-Armed Bandits to Conduct Adaptive Educational Experiments, Journal of Educational Data Mining, 11(1), 47-79.
5. White, J. M. (2012), Bandit Algorithms for Website Optimization, O'REILLY.
6. Clement, B., Roy, D., Oudeyer, P-Y, Lopes, M., (2015), Multi-Armed Bandits for Intelligent Tutoring Systems, Journal of Educational Data Mining, 7(2), 20-48 (2015)
7. Manickam, I., A. S. Lan and R. G. Baraniuk, "Contextual multi-armed bandit algorithms for personalized learning action selection," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, 2017, pp. 6344-6348, doi: 10.1109/ICASSP.2017.7953377.
8. Melesko, J.; Novickij, V. Computer Adaptive Testing Using Upper-Confidence Bound Algorithm for Formative Assessment. *Appl. Sci.* **2019**, *9*, 4303.
9. Agrawal, R., Sample mean based index policies by *O*(log n) regret for the multi-armed bandit problem. Advances in Applied Probability, 1995, 27(4). 1054-1078.
10. Thompson, W. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Biometrika, 25(3/4): 285-294
11. Chapelle, O. and Li, L. (2011), An empirical evaluation of Thompson sampling. In: Advances in Neural Information Processing Systems, 24. 2249-2257
12. Graepel, T., Candela, J. Q., Borchert, T., Herbrich, R., Web-Scale Bayesian Click-Through Rate Prediction for Sponsored Search Advertising in Microsoft's Bing Search Engine, In: Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa, Israel
13. Hill, D. N., Nassif, H., Liu, Y., Iyer, A., and Vishwanathan, S. V. N.. 2017. An Efficient Bandit Algorithm for Realtime Multivariate Optimization. In Proceedings of KDD'17, Halifax, NS, Canada, pp. 1813-1821, 2017