Personalized Exams and Learning in Massive Open Online Courses

Pushkar Kolhe PUSHKAR@CC.GATECH.EDU
Georgia Institute of Technology, 810 Atlantic Drive., Atlanta, GA 30332 USA

Dr. Michael L. Littman MLITTMAN@CS.BROWN.EDU
Brown University, 115 Waterman St. Providence, RI 02912, USA

Dr. Charles I. Isbell ISBELL@CC.GATECH.EDU
Georgia Institute of Technology, 810 Atlantic Drive., Atlanta, GA 30332 USA

Abstract

Personalized education is becoming an important aspect of Massive Open Online Courses. A part of that is to provide personalized content to every student. In this paper we design a methodology to formulate personalized exams. As a consequence of this formulation we are also able to define a metric for learning: an improvement in the students proficiency because of the course. In our present education system professors are rewarded as a parameters of the number of students who score good in an exam and the student is rewarded for a good score. The optimal action for the teacher is to set an easy exam so that the student scores well. We formulate an exam as a bayesian game and use auction theory to create motivation for the teacher to help the student learn.

1. Introduction

The role of personalized education has become more important with the rise of Massive Open Online Courses (MOOC). In traditional classroom settings standardized tests are used to test how much a student learnt in this class. The goal of personalized education is to set up a learning environment where the content and exams are personalized to the ability of a student. Personalization can also help in setting up personal goals that motivate the student to learn at their own pace. In this paper we postulate a framework with game theory and auction theory that does exactly that.

The dropout rate in MOOC courses is around 85-90% (Rivard, 2013). We think one of the main reasons behind this is that the course are not yet personalized. In a regular university setting all students are equally skilled and the course is designed for them. Online the skill set of students is quite varied. Personalization is even more important. There has been some work in creating personalized content (Henning et al., 2014; Paquette et al.). In this paper we show how personalized exams can be created based on the proficiency of a student. We will also show that we can use the assessment from these personalized exams to track a students progress through a course and measure their progress quantitatively.

Some MOOC providers also provide certification to their students. To keep a high rate of completion the content provided by them is usually set so that the goals needed to complete the course are easier than they should be. If there was a way to set personalized goals, perhaps, the providers would be motivated to provide content for a wide variety of students.

In a classical exam every student's progress is evaluated with the same exam. The students grade depends on how well he scored in this exam. If a student answers all questions correctly he has done well in the class. However it is hard to track if he actually learnt anything. Every student starts out a course with different background and skill set. We have found out from our experience in teaching a class that a student's grade is best correlated with the progress a student makes in the class. In that sense we propose an approach where we set up a framework where we can measure progress of a student by giving them personalized exams.

2. Our Approach

We assume that the student has to take a set number of exams throughout the course. The student makes as much progress as they can. The framework has two parameters: proficiency of the students and difficulty of questions. They can make progress through the course by attempting hard questions. We have found in our classes that students learn

Proceedings of the 32^{nd} *International Conference on Machine Learning*, Lille, France, 2015. JMLR: W&CP volume 37.

when they answer questions that they previously found difficult. In this approach we show how we could formulate an exam using game theory and measure the number of questions a student's answers that were previously difficult.

2.1. Item Response Theory

The first question is to initialize all the students with some proficiency and questions with difficulty. To calculate the proficiency and hardness of the questions we use the Item Response Theory as proposed by (Naeff & Nichols, 2014). The most intuitive way to find out how proficient a student is at a certain concept is to compute what percentage of questions they answered correctly. However that doesnt take into consideration the hardness of the questions. In fact if we want to compute how hard a question was we can compute it based on the students with a certain proficiency who answered it correctly. To model this relationship between student proficiency and question hardness, Naeff et. al. use a sigmoid function given by,

$$P(correct) = \frac{1}{1 + e^{-(\theta - \beta)}} \tag{1}$$

Here θ represents proficiency of a student and β represents the questions hardness. We perform the Item Response Theory (IRT) fitting process by maximizing the log-likelihood given the dataset. Additional regularization is introduced by Bayesian priors, so that a priori, difficulties β and abilities θ are assumed to be normally distributed according to $N(0, 1)$. This yields the objective function:

$$Z = -\sum_{(s,i)\in R} (\log(1 + e^{-\theta_s - \beta_i}) + (1 - r_{s,i})(\theta_s - \beta_i)) - \sum_{s} \frac{1}{2}\theta_s^2 - \sum_{i} \frac{1}{2}\beta_i^2, \tag{2}$$

where $r_{s,i}$ is the response of student s to question i, where a correct response is coded as 1 and an incorrect response is coded as 0. All the responses are stored in R, which contains a pair (s, i) where the students have given a response. To predict individual student responses with IRT, we use the student and item parameters found via maximizing (2). Following (1) we predict that student s will answer question i correctly only if $\theta_s \geq \beta_i$.

2.2. Personalized Exams

Let us consider this classical assessment in the context of a game theoretic framework where the exam is a game between the teacher and the student. Here the student is happy when he scores an A grade. The teacher is happy when all of their students score on A grade. In this framework the Nash equilibrium would motivate the teacher to set an easy exam for the student.

	Correct	Incorrect
$Easy$	1	-3
$Hard$	3	-1

Figure 1. An Example payoff matrix. The student and the teacher get the same payoff. The teacher can choose an easy or hard question, whereas the student can answer it correctly with probability p or incorrectly with probability $1 - p$.

In a personalized assessment a different exam is created for every student based on their proficiency. Let us consider this exam in the context of a game theoretic framework between the teacher and the student. In this game the student and the teacher is rewarded every time the student answers a question that the teacher thinks is beyond the student's proficiency. The teacher has to strategically pick questions that are below or slightly above the students proficiency since there is a high probability that the student will answer them.

2.2.1. Exam as a Bayesian Game

Consider the above payoff matrix. If the difficulty of a question is greater than the students proficiency than the question is hard, otherwise it is easy. The probability that a student answers the hard question correctly is less than for the easy question. The magnitude of this probability function depends on the students proficiency.

If the teacher asks an easy question and the student answers it correctly they get a point, but they get 3 points to answer a hard question. On the other hand they lose more points for incorrectly answering an easy question than a harder one. In the end of a personalized exam, the students proficiency is recalculated based on IRT.

If we postulate that the teacher wants the student to learn quickly the teachers job is to find out the best questions to ask. If the teacher asks the hardest questions there is a chance that the student answers the question incorrectly. So the teacher has to choose questions so that the expected payoff is maximum.

In fact a generalization can be viewed as a Bayesian Games where the teacher has to pick the best strategies, best questions in this case, so that the expected payoff is maximum. If an exam consists of n questions, the teacher has to choose the best n strategies that will increases their expected payoff.

	Correct	Incorrect	Payoff
2(4.0)	3.69	−3.69	2.35
3(4.3)	4.29	−4.29	2.31
5(5.8)	9.09	−9.09	−1.35
6(6.0)	10.04	−10.04	−2.46
7(6.9)	15.75	−15.75	−9.52

Figure 2. Example payoff matrix for a simulation we ran for a course. The student proficiency is 5.5. The number in the bracket in the left column correspond to the difficulty of the question as predicted by IRT. The third column is not a part of the payoff matrix. It is the computed expected payoff for each question. You can see that the questions with negative expected payoff have difficulty greater than the student proficiency.

2.2.2. Payoff Values of the Game

The payoff values for every question are determined by the difficulty parameter from IRT. The student can solve a question whose difficulty is equal to the students proficiency. The payoff should be low for answering questions below their proficiency and the payoff should be high for answering a question that is really difficult. So we use the exponential distribution for calculating the payoff. The payoff is given by,

$$payoff_i = -\lambda e^{-\lambda\beta_i} \tag{3}$$

where $payoff_i$ is the payoff for answering question i with difficulty β_i correctly. The payoff is negative if the student answers the question incorrectly.

To find the Bayes Nash equilibrium strategy the teacher has to pick up questions that have the best expected payoff. The probability that a student with proficiency θ answers the question correctly is given by (1). (3) and (1) can be used to calculated the expected payoff.

2.2.3. Mechanism Design

Assumption 1 *The question bank consists of an infinitely many questions of each difficulty level.*

For the purposes of generalization, we made the assumption that the question bank consists of unlimited questions of every difficulty level. If we setup a Bayesian game with this assumption, the teacher will never pick questions that are harder than the proficiency of the student. To motivate the teacher to pick hard questions we set up rewards we look at this game as a Vickrey auction (MacKie-Mason & Varian, 1994; Ausubel & Milgrom, 2006).

To understand why we need to do this, consider that we have set up exams as a bayesian game. The teacher is trying to maximize their payoff. There is no rational motivation to pose a question where the expected payoff is negative.

Consider that the teacher is trying to bid for predicting the value of a students proficiency. The other bidder is the teacher bidding at the next exam. Like Vickrey auction we set up the game so that the teacher does not lose any payoff for the hardest question they picked for an exam. But if the student answers the question correctly, they get the payoff.

This motivates the teacher to ask at least one question where the expected payoff is negative, in the hope that if the student answers the question correctly they will get a huge payoff and if the student answers it incorrectly they dont lose anything. If the student answers the hard question correctly, their proficiency can increase.

We conjecture that if you measure the number of times student answer this hard question correctly is a good measure of learning. Consider the case that the student never answers this one question correctly. The student came into a course with some proficiency. If they never learnt anything in the class they would never increase their proficiency score.

Conjecture 1 *Students learn when they can answer a hard question that their teachers did not expect the student to answer based on their proficiency.*

3. Conclusion

We created personalized games on a simulated dataset. Since we made the assumption that we have infinitely many questions of the same difficulty level, we had multiple answers for a pair of student and question. If we did not make this assumption then after a few exams the teacher chooses questions whose expected payoff is negative, that is questions harder than the proficiency of the student. The auction strategy as described in Section 2.2.3 is redundant.

In this paper we only described an exam that is a game between the teacher and a student. The learning metric cannot be used for grading. If it is used for grading there is a chance that the student takes the initial exam so that he gets a low proficiency. In the subsequent exams the hardest question is easy for this student and he would easily get a good grade. In this paper we assume that the student is a rational agent who is trying to maximize their payoff. If we had multiple students, as in university courses, the grade should likely depend on other factors such as: the number of exams it took to attempt the hardest exam, the proficiency after a limited exams and so on.

References

Ausubel, Lawrence M and Milgrom, Paul. The lovely but lonely vickrey auction. *Combinatorial auctions*, 17:22–26, 2006.

Henning, Peter A, Heberle, Florian, Streicher, Alexander, Zielinski, Andrea, Swertz, Christian, Bock, Jürgen, Zander, Stefan, et al. Personalized web learning: Merging open educational resources into adaptive courses for higher education. *Personalization Approaches in Learning Environments*, 55, 2014.

MacKie-Mason, Jeffrey K and Varian, Hal R. Generalized vickrey auctions. 1994.

Naeff, Ruben and Nichols, Zachary. On the limits of psychometric testing in online education. 2014.

Paquette, Gilbert, Marino, Olga, Rogozan, Delia, and Léonard, Michel. Competency-based personalization for massive online learning.

Rivard, Ry. Measuring the mooc dropout rate. *Inside Higher Ed*, 8:2013, 2013.