Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)

Incentivizing Peer Grading in MOOCS: An Audit Game Approach

Alejandro Carbonara and **Anupam Datta**
Carnegie-Mellon University
auc,danupam@cs.cmu.edu

Arunesh Sinha
University of Southern California
aruneshs@usc.edu

Yair Zick
Carnegie-Mellon University
yairzick@cmu.edu

Abstract

In Massively Open Online Courses (MOOCs) TA resources are limited; most MOOCs use peer assessments to grade assignments. Students have to divide up their time between working on their own homework and grading others. If there is no risk of being caught and penalized, students have no reason to spend any time grading others. Course staff want to incentivize students to balance their time between course work and peer grading. They may do so by auditing students, ensuring that they perform grading correctly. One would not want students to invest too much time on peer grading, as this would result in poor course performance. We present the first model of strategic auditing in peer grading, modeling the student's choice of effort in response to a grader's audit levels as a Stackelberg game with multiple followers. We demonstrate that computing the equilibrium for this game is computationally hard. We then provide a PTAS in order to compute an approximate solution to the problem of allocating audit levels. However, we show that this allocation does not necessarily maximize social welfare; in fact, there exist settings where course auditor utility is arbitrarily far from optimal under an approximately optimal allocation. To circumvent this issue, we present a natural condition that guarantees that approximately optimal TA allocations guarantee approximately optimal welfare for the course auditors.

1 Introduction

Massive Open Online Courses (MOOCs), such as those offered by Coursera, are online courses open to the general public. Those who enroll in such a course are usually required to complete ongoing assignments as part of the course requirements. However, as their name suggests, MOOCs tend to have student enrollment in the thousands, with only a limited number of teaching assistants available to grade assignments. To overcome this issue, MOOCs employ *peer assessment*: the students grade each other, with the course staff acting as high-level moderators. There are several obvious issues with peer grading. First, students' ability to grade is likely to be quite heterogeneous: some students naturally have a better grasp of course material than others. Second, students may not be very commited to grading their peers. While Coursera does impose a 20% grade penalty on students who fail to complete their grading assignments on time, this says nothing about the quality of the work. Though there are educational merits to grading others' work (being exposed to different solutions, verifying that one's solution was correct and so on), even the most hard working student may not be interested in spending time on grading if that time can be better spent on completing their own homework assignments.

That said, if course staff employ an overly strict policy towards students who poorly grade their peers, they risk poor student performance on class assignments. Students will spend too much time on peer grading and too little on the actual assignment. To further complicate matters, students who feel that staff is treating them unfairly will simply drop out. Course staff control students' time division by controlling the likelihood that a student's grading is audited. A high likelihood of audit will lead to a higher likelihood of being caught misgrading which in turn leads to the student investing more into peer grading over working on the assignment. Since the hours one can spend auditing are limited, one must carefully assign TAs to student graders in order to ensure that they behave in a desirable manner. This leads to the following natural question: *what assignment of TA auditors to students best incentivizes them to optimally balance their time between grading and completing assignments?*

In this work, we provide some insights into this matter.

1.1 Our Contribution

We model this setting as a non-cooperative game between course staff and students; the solution concept we employ is a Stackelberg equilibrium. Both choices are natural in this setting.

First, both students and course staff have actions that they may take to further their agendas: students' actions are a choice of the number of hours to assign to grading vs. the number of hours to assign to completing the assignment. The course staff/auditors' actions are an allocation of TAs/auditors to students. Furthermore, both course staff and students have natural utility functions; in the case of students their utility is their grade, and in the case of the course staff their utility is the overall class grade minus a penalty for each misgrading that occurs. A misgrading penalty may encode the number of hours that are invested in dealing with student complaints or expected dropout rates.

Next, Stackelberg equilibrium is an equilibrium concept where one side gets to observe the (mixed) strategy of another. The first player who moves is called the *defender* and the player who gets to observe the defender's strategy is called the *attacker*. In our setting we think of the course staff as the defenders and the students as the attackers. Thus, the course staff publicly declare how they plan to audit students (perhaps via their publicly available course handbook). The students must choose how much effort they invest in their homework assignments and how much they invest in grading.

Note that a Stackelberg security game [Tambe, 2011] traditionally has one attacker and one defender, whereas our setting handles multiple independent non-uniform attackers.

We begin by showing that finding a TA allocation that maximizes the expected utility of the course staff is generally NP-hard. However, there exists an efficient dynamic programming algorithm that is able to find a TA allocation ε close to the defender-optimal allocation. This solution is, unfortunately, unsatisfactory: we show that there exist games where a TA allocation ε close to the optimal allocation does not guarantee that the defender's utility is ε-close to optimal. In fact, it is possible that a TA allocation that is close to the optimal but has attacker best-responses that offer an arbitrarily low utility to the defender. Intuitively, this may happen since students who have only slightly less TA resources allocated to them may decide to switch from investing a lot in peer grading, to investing nothing. However, we show that under some mild assumptions on the students' utility functions, we can guarantee that TA allocations that are ε-close to the optimal guarantee ε-optimal utility to the defenders.

We assume that the course staff have full information of students' utilities. While this is a rather strong assumption, it can be justified. As Piech *et al.* [2013] show, it is possible to elicit student grading abilities and expected class performance based on their past grading performance.

1.2 Related work

The most well-known application of Stackelberg games is security games [Tambe, 2011; Jain *et al.*, 2010; Kiekintveld *et al.*, 2009; An *et al.*, 2011]. A key difference between our work and the standard security game model is that we assume multiple attackers with different utility functions. While previous applications of Stackelberg games for physical security have considered multiple attacks by an attacker [Korzhyk *et al.*, 2011], we separate ourselves from them by giving each attacker a disjoint objective. In our model, defenders want to punish attackers for misbehaving, but do not want to do so too much. This assumption is employed in audit games [Blocki *et al.*, 2013; 2015]. However, Blocki *et al.* [2013] and Blocki *et al.* [2015] assume that they are given some parameter α representing the "punishment level" for the defender. In our setting, over-auditing of certain students may cause them to invest too much effort into grading, which may not be optimal for the course staff. In other words, in our setting the motivation not to over-audit is *intrinsic* rather than a given parameter.

The problem of handling low quality peer grading has been studied from various aspects. Piech *et al.* [2013] focus on using machine learning techniques on peer grading networks in order to find the true scores of homework assignments. Wright *et al.* [2015] design and test a peer grading mechanism suitable for small sized courses. Their idea is to initially let the TA check every student's grading and then promote some students to an independent grader level, while adding the provision that students may complain about their grades. Shalem *et al.* [2014] design a graphical model based approach to predict student performance in a MOOC. Ghosh and Kleinberg [2013] focus on designing incentive mechanisms to increase student participation in the course discussion forum. Vallam *et al.* [2015] model student participation in discussion forums as a game; Li and Conitzer [2013] use a Stackelberg game model to set exam questions. Finally, Jurca and Faltings [2003] propose an incentive compatible peer reputation mechanism; however, their setup is quite different from ours.

2 Preliminaries

Let $N = \{1, \dots, n\}$ be the set of *students*. Each student $i \in N$ needs to decide how much time to allocate to homework, and how much to allocate to peer grading. Thus, a student's strategy is a value $x_i \in [0, 1]$, where x_i is the percentage of time devoted to homework. The *instructor* has a total of k_{tot} resources. It is convenient to think of k_{tot} as the number of TA-hours available to the instructor. The instructor's allocation is a vector $\mathbf{k} \in \mathbb{R}^n_{\geq 0}$ such that $\sum_{i=1}^n k_i \leq k_{tot}$, where k_i is the probability that student i is audited.

Observe that specifying $\mathbf{k}$ is enough to recover a probability distribution over pure allocations of TAs to students — i.e. an instructor mixed strategy — by the Birkhoff von-Neumann decomposition (see [Korzhyk *et al.*, 2011] for details, and [Birkhoff, 1946] for the original result).

The game proceeds as follows. First, the instructor chooses a vector $\mathbf{k}$ such that $\sum_{i \in N} k_i \leq k_{tot}$. Each student $i \in N$ observes $\mathbf{k}$, and chooses $x_i \in [0, 1]$ that maximizes her utility. We assume that every student's utility is completely identifiable with her grades on the assignment. More formally, a student's utility function is given by two real-valued functions: H_i and M_i. $H_i(x)$ is the expected grade of i if she invests an x portion of her time in homework, and $M_i(1-x)$ is the expected probability that she is punished for misgrading her peers if audited assuming she invests $(1-x)$ percent of her time on grading. The punishment is scaling, as we do not want to have a case where the punishment is greater than the student's entire homework score. Putting it all together, we get that the expected utility of student i under $\mathbf{k}$ if she invests an x_i portion of her time in homework is

$$U_i(x_i, k_i) = H_i(x_i)\left(1 - k_i M_i(1 - x_i)\right). \qquad (1)$$

We assume that H_i is monotone increasing; similarly, we assume that M_i is monotone decreasing. We note that there still may be a chance of misgrading even with a high amount of effort. That is, we are not guaranteed that $M_i(1) = 0$.

Next, let us consider the utility of the instructor. The instructor is interested in getting students to put reasonable effort into both their homework assignment and into peer grading. The utility that the instructor gets from homework is simply a weighted average over students. That is, each student is associated with a weight α_i, and the instructor's utility from homework if students' strategies are given by $\mathbf{x}$ is given by $\sum_{i=1}^{n} \alpha_i H_i(x_i)$. Similarly, the instructor incurs a penalty β_i for each student i that has misgraded; thus, the instructor's expected penalty given $\mathbf{x}$ is $\sum_{i=1}^{n} \beta_i M_i(1-x_i)$. These different values of α_i and β_i represent a weighted average over the different types of student. For example, undergraduate students may be expected to perform better on homework, masters students may be expected to focus more on grading, and students who are auditing the class may be expected to spend less time overall. These values are determined solely by the instructor's preferences, and are not influenced by student behavior. Putting it all together, the instructor's utility function is given by

$$U^{INST}(x_1, \dots x_n) = \sum_{i=1}^{n} (\alpha_i H_i(x_i) - \beta_i M_i(1-x_i)) \quad (2)$$

We are interested in finding a TA allocation $\mathbf{k}$ that maximizes the instructor's utility, assuming that students observe $\mathbf{k}$ and then play a best response to their level of audit.

Definition 2.1 (Instructor-Optimal Stackelberg Equilibrium). An instance of INSTRUCTOROPTEQ is given by:

- A number of students N
- For each student a homework and misgrading function $H_i, M_i : [0,1] \to [0,1]$
- For each student an instructor utility function weighted by $(\alpha_1, \dots, \alpha_n); (\beta_1, \dots, \beta_n)$
- Audit resources given by k_{tot}
- Target value V

It is a "yes" instance if there exists some instructor mixed strategy $\mathbf{k}$ such that for students' best response to $\mathbf{k}$, the instructor's utility is at least V; it is a "no" instance otherwise.

The optimization variant of Definition 2.1 is

$$\begin{aligned}
&\max \sum_{i=1}^{n} \alpha_i H_i(x_i) - \beta_i M_i(1-x_i) && (3)\\
&\text{s.t. } x_i \in \arg\max_{x \in [0,1]} U_i(x, k_i) && \forall i \in N\\
&\qquad \sum_{i=1}^{n} k_i \le k_{tot}\\
&\qquad k_i \in [0,1] && \forall i \in N.
\end{aligned}$$

3 Computing Instructor-Optimal Strategies

We begin by showing that INSTRUCTOROPTEQ is computationally intractable for simple step-shaped homework and misgrading functions.

Theorem 3.1. INSTRUCTOROPTEQ *is NP-hard.*

Proof. Our reduction is from the KNAPSACK problem [Garey and Johnson, 1979]. An instance of KNAPSACK is given by a list of n objects $N = \{1, \dots, n\}$, each with a positive weight w_i and positive value v_i, a value $V \in \mathbb{R}_+$ and a weight $W \in \mathbb{R}_+$. It is a "yes" instance if there is some $S \subseteq N$ such that $\sum_{i \in S} w_i \le W$ and $\sum_{i \in S} v_i \ge V$, and a "no" instance otherwise. Given an instance of KNAPSACK $\langle N = \{1, \dots, n\}, \mathbf{w}, \mathbf{v}, V, R \rangle$, We assume that v_i and w_i are positive integers for all $i \in N$; furthermore, we assume that $W \ge w_i$ for all $i \in N$, and that $V < v(N) = \sum_{i=1}^{n} v_i$.

We construct the following game: The items are the students. For each student $i \in N$ we define

$$H_i(x) = \begin{cases} \frac{v_i}{C} & \text{if } x = 1 \\ 0 & \text{otherwise,} \end{cases}$$

where $C > 0$ is some very large constant. We also write

$$M_i(x) = \begin{cases} \frac{W}{w_i C} & \text{if } x = 0 \\ 0 & \text{otherwise.} \end{cases}$$

We set $\alpha_i = 0$ and $\beta_i = w_i$, making the instructor utility $-\sum_{i \in N} w_i v_i M_i(1-x_i)$. That is, the instructor does not care at all about students' performance in the course, just about minimizing their misgrading probability, weighted by $w_i v_i$. We assume that the staff has $k_{tot} = 1$, and our student utility is $-\frac{W}{C}(v(N) - V)$. Suppose that student i is audited with probability k_i and she assigns all of her efforts to doing her homework; then $U_i(1, k_i) = \frac{1}{C} v_i (1 - k_i \frac{W}{w_i})$. For any $x < 1$, $U_i(x, p_i) = 0$. Therefore, the best response of i to k_i is $x_i = 1$ if $k_i < \frac{w_i}{W}$, and is $x_i < 1$ otherwise.

Suppose that we have a "yes" instance of knapsack. This means that there is some set of students S^* such that $\sum_{i \in S^*} w_i \le W$ and their total value is $\ge V$. Suppose that we set $k_i^* = \frac{w_i}{W}$ for all $i \in S^*$ and $k_i^* = 0$ for all $i \in N \setminus S^*$. Since $\sum_{i \in S^*} w_i \le W$ this is a valid instructor strategy. Under this strategy, instructor utility is

$$\begin{aligned}
-\sum_{i=1}^{n} w_i v_i M_i(1-x_i) &= -\frac{W}{C} \sum_{i \in N \setminus S^*} w_i \frac{v_i}{w_i} \\
&= -\frac{W}{C} \sum_{i \in N \setminus S^*} v_i \\
&\ge -\frac{W}{C}(v(N) - V).
\end{aligned}$$

On the other hand, if there exists an audit strategy $\mathbf{k}^*$ that guarantees utility of at least $-\frac{W}{C}(v(N) - V)$ we observe that setting $k_i^* > \frac{w_i}{W}$ to student i adds no benefit to the instructor, and if $k_i^* < \frac{w_i}{W}$ one may as well set $k_i^* = 0$. Thus, we assume w.l.o.g. that $k_i^* \in \{0, \frac{w_i}{W}\}$ for all $i \in N$. Then, the students audited with probability $\frac{w_i}{W}$ correspond to a subset of items that weigh less than W and have a value of at least V, which concludes the proof. □

4 Finding Approximately Optimal Strategies

Since computing the optimal solution is computationally infeasible, we then consider finding an approximate solution. If we limit the precision with which we express the probabilities k_t, then we obtain an efficient dynamic programming algorithm for solving the instructor-optimal Stackelberg equilibrium problem. Our algorithm involves solving sub-problems where we only consider a subset of our students with a subset of our resources. We define $x_j(k)$ to represent the maximum possible utility our instructor can get from all the students with indices $t \leq j$ with k amount of audit resource to allocate. Note that this k is not necessarily a whole number. Additionally, note that if $k > j$, $x_j(k) = x_j(j)$ as at most there is one resource allocated to each student.

Furthermore, define $f_t(k)$ to be the utility the instructor gets from just the student t when k audit resources are allocated to him. In general, computing $f_t(k)$ is dependent on the student's utility functions. Let $||f||$ be the worst case evaluation time of this utility response function.

Suppose we allocate our auditors with maximum bit precision b. From this, we propose Algorithm 1:

Algorithm 1: Solving Fixed Precision K

```
Data: b, k_tot, T = {1, 2, ... n}, f_t(k)
Result: x_t(k_tot)
∀_{t,k} : x_t(k) = −∞;
for j = 1; j ≤ n; j++ do
    for k = 0; k ≤ k_tot; k = k + 2^{−b} do
        if j = 1 then
            x_j(k) = max(x_j(0), x_j(k − 2^{−b}), f_j(k));
        else
            for l = 0; l ≤ 1; l = l + 2^{−b} do
                x_j(k) =
                max(x_{j−1}(k − l) + f_j(l), x_j(k));
```

Theorem 4.1. *If we are limited to specifying our audit strategy with b bits of precision, Algorithm 1 finds the instructor-optimal equilibrium with running time in $O(k_{tot}4^b n||f||)$.*

Proof. First, we reason about the correctness of the algorithm. Suppose for the current iteration of the loop in the algorithm, the indices are j and k. We can show that for all integral j' and for all k' with b bits of precision, we will have already found the optimal solution to the subproblem $x_{j'}(k')$. We do so using induction on j.

Our base case where $j = 1$ computes the following. We iterate through all possible values of $f_1(l)$ and take the one that gives the most utility.

For any later loops, inductively assume that we have solved the subproblem for all values $j' < j$, and for all possible values of k. Our optimal solution will assign at most 1 unit of audit to the jth target. We consider every possible $f_j(l) + x_{j-1}(k-l)$. Since our total utility is equal to the sum of the utility from each target, the maximum of this value will be the maximum possible utility for $x_j(k)$. This completes our claim about optimality in each iteration j.

Thus, at the end of our algorithm we get the maximum possible utility for $x_n(k_{tot})$, and therefore we will have the maximum possible utility our instructor can achieve with the resources available to them.

Regarding the running time: note that we have three levels of loops. The inner loop will iterate at most 2^b times, the second level loop will iterate $k_{tot}2^b$ times, and the topmost will iterate at most n times. At the core of each loop, we will need to compute $f_j(k-l)$ and compare several values. Recall that computing $f_j(k-l)$ takes worst case $||f||$ time. Therefore, our algorithm will take time $O(k_{tot}4^b n||f||)$. □

While the above algorithm is efficient for a fixed b (or for a b proportional to $log(k_{tot})$), the important question is whether we can guarantee that the discrete rounded solution provides utility close to the optimal solution. In section 4.2, we identify the conditions under which we can provide ε'-approximation guarantees for any constant ε'. In the next section, we demonstrate that guaranteeing an approximate best response from the students could potentially result in a non-negligible loss of utility for the instructor.

4.1 ε-Best Response

Suppose we are given an optimal (unrounded) solution k_i and the students' responses are x_i. Suppose we evaluate an approximate strategy k_i* that is within ε ($\varepsilon = 2^{-b}$) of the true solution. Additionally, suppose that the student's utility functions are Lipschitz continuous, which means that the function's rate of change is limited. We can then show that x_i is an ε-best response to k_i*, meaning that we can cap the loss of utility that a student has for playing x_i instead of whatever is the true optimal response to k_i^*. In order to show this, we first prove the following lemma.

Lemma 4.2. *Let $f : [0,1]^2 \to \mathbb{R}$ be a function such that for any fixed x_0 and y_0, $f(x, y_0)$ and $f(x_0, y)$ are Lipschitz continuous with constant c. Then the function*

$$F(y) = \max_{x \in [0,1]} f(x, y)$$

is Lipschitz continuous with constant c, i.e.,

$$|F(y_1) - F(y_2)| < c|y_1 - y_2|.$$

Proof. Let $y_1, y_2 \in [0,1]$ be two points; let x_1^*, x_2^* be two points such that $x_1^* \in \text{argmax}_{x \in [0,1]} f(x, y_1)$ and $x_2^* \in \text{argmax}_{x \in [0,1]} f(x, y_2)$. As these function are Lipschitz continuous over a compact set, the max values are within the compact set. Then $F(y_1) = f(x_1^*, y_1)$ and $F(y_2) = f(x_2^*, y_2)$. Without loss of generality, suppose that $F(y_1) \geq F(y_2)$. By Lipschitz continuity, $|f(x_1^*, y_1) - f(x_1^*, y_2)| \leq c|y_1 - y_2|$.

Now, either $f(x_1^*, y_1) \geq f(x_1^*, y_2)$ or $f(x_1^*, y_1) < f(x_1^*, y_2)$. If $f(x_1^*, y_1) < f(x_1^*, y_2)$, then using the fact that $f(x_1^*, y_2) \leq f(x_2^*, y_2)$ we can conclude $f(x_1^*, y_1) < f(x_2^*, y_2)$. The last inequality is nothing but $F(y_1) < F(y_2)$, a contradiction; therefore $f(x_1^*, y_1) \geq f(x_1^*, y_2)$.

Next, by Lipschitz continuity we know that

$$f(x_1^*, y_1) \leq c|y_1 - y_2| + f(x_1^*, y_2),$$

using the fact that $f(x_1^*, y_2) \leq f(x_2^*, y_2)$. Combining the above two inequalities we conclude that $f(x_1^*, y_1) \leq c|y_1 - y_2| + f(x_2^*, y_2)$, which is same as $|F(y_1) - F(y_2)| \leq c|y_1 - y_2|$. Thus, $F(y)$ is Lipschitz continuous. □

We are now ready to prove the following result:

Theorem 4.3. *Suppose that H_i and M_i are Lipschitz continuous with constant c. Suppose that k_i is the instructor-optimal solution. Let x_i be the student's best response to an audit level of k_i, and choose k_i^* such that $|k_i - k_i^*| \le \varepsilon$.*

Then x_i is a $c\varepsilon$-best response to k_i^*

Proof. We have that H_i and M_i are Lipschitz continuous. Let $OPT(k_i)$ be the best response of student i to an audit level of k_i. According to Equation 1, student utility is given by

$$U_i(x, k_i) = H_i(x)(1 - k_i M_i(1 - x)).$$

In an equilibrium, students will play their best response strategies in response to k_i. We thus define the utility of a student playing their optimal strategy as

$$OPT(k_i) = \max_{x \in [0,1]} H_i(x)(1 - k_i M_i(1 - x)).$$

We have that both H_i and M_i are Lipschitz continuous. Therefore, for a fixed k_0 we have $UTIL(x, k_0)$ is Lipschitz continuous. Suppose we are given y_1, y_2 such that $|y_1 - y_2| < \varepsilon$. Since both H_i and M_i are positive and bounded by 1 we have that for any fixed x_0

$$|U_i(x_0, y_1) - U_i(x_0, y_2)| < \varepsilon.$$

Thus, applying Lemma 4.2 we claim that $OPT(k_i)$ is Lipschitz continuous, i.e.,

$$|OPT(y_1) - OPT(y_2)| < c|y_1 - y_2|.$$

Therefore, if $|k_i - k_i^*| \le \varepsilon$, then $|OPT(k_i) - OPT(k_i^*)| \le c\varepsilon$. □

4.2 Exact Best Response

However, even if x_i is an $c\varepsilon$-best response to k_i^*, what could be said about the exact best response x_i^* to k_i^*? We would hope that this response x_i^* would be close to x_i. Unfortunately, that need not be true for an arbitrary Lipschitz continuous function, as we show in the following example.

Example 4.4. Define our utility functions to be:

$$H_i(x_t) = \begin{cases} \frac{1}{10} & \text{if } x_i \ge \frac{1}{2} \\ \frac{1}{10} + \frac{1}{5}(\frac{1}{2} - x_i) & \text{if } x_i < \frac{1}{2} \end{cases}$$

$$M(x_t) = \begin{cases} \frac{9}{10} & \text{if } x_i > \frac{1}{2} \\ \frac{9}{10} - \frac{1}{5}(\frac{1}{2} - x_i) & \text{if } x_i \le \frac{1}{2} \end{cases}$$

If we then plug these values into our utility function:

$$\begin{aligned} &H_i(x_i)(1 - k_i M_i(1 - x_i)) \\ &= \begin{cases} (\frac{1}{10} + \frac{1}{5}(\frac{1}{2} - x_i))(1 - \frac{k_i 9}{10}) & \text{if } x_i < \frac{1}{2} \\ (\frac{1}{10})(1 - k_i(\frac{9}{10} - \frac{1}{5}(x_i - \frac{1}{2}))) & \text{if } x_i \ge \frac{1}{2} \end{cases} \end{aligned}$$

If we consider the two cases disjointly, the two strategies that the student would consider are $x_i = 0$ and $x_i = 1$. The payoffs the student will get from those strategies are $\frac{9}{10}(1 - \frac{9}{10}k_i)$ and $\frac{1}{10}(1 - \frac{1}{10}k_i)$ respectively.

If $k_i = 1$, we can see that the student could potentially choose 1. However, if k_i is any less, the student would opt for 0. Remember that the utility function of our instructor is of the form:

$$\sum_i (\alpha_i H_i(x_i) - \beta_i M_i(1 - x_i))$$

If we then set $\alpha_i = 0$, we get that this represents a significant drop in the instructor's utility.

Next we characterize the cases where Algorithm 1 guarantees the instructor utility to be close to the optimal. In order to do this, we first present conditions under which the student's strategy is stable under small changes to k_i.

Theorem 4.5. *Let k_i and k_i^* be two different possible audit levels such that $|k_i - k_i^*| < \varepsilon$. Let x_i and x_i^* be the response of our student under k_i and k_i^* respectively.*

Suppose the student's utility functions $\forall k_0 : f(x, k_0)$ is strictly concave, $\forall k_0 : f''(x, k_0) < -d$, and $\forall x_0, k_0 : f(x, k_0), f(x_0, k)$ are Lipschitz continuous.

We then have $|x_i - x_i^*| < \sqrt{(4c\varepsilon/d)}$.

Proof. We are given k_i and k_i^*. By assumption, x_i (x_i^*) is the maximizer of $f(x, k_i)$ ($f(x, k_i^*)$). Therefore $f(x_i, k_i) > f(x_i^*, k_i)$ and similarly $f(x_i, k_i^*) < f(x_i^*, k_i^*)$.

Remember that $|k_i - k_i^*| < \varepsilon$. Due to the Lipschitz continuity of $f(x_0, k)$:

$$\forall x_0.\ |f(x_0, k_i) - f(x_0, k_i^*)| < c\varepsilon$$

For any positive v, consider $x_i + v$. Recall that: $\forall k_0 : f''(x, k_0) < -d$. The Taylor expansion of $f(x_i + v, k_i)$ (in remainder form) is given by

$$f(x_i, k_i) + v f'(x_i, k_i) + \frac{v^2 f''(\hat{x}, k_i)}{2}.$$

where $\hat{x}$ lies between $x_i + v$ and x_i. Since $f'(x_i, k_i)$ is zero and $f''(\hat{x}, k_i) < -d$ we get

$$f(x_i + v, k_i) + \frac{dv^2}{2} < f(x_i, k_i).$$

We can use a similar derivation to get

$$f(x_i - v, k_i) + \frac{dv^2}{2} < f(x_i, k_i).$$

Next, for contradiction, we suppose that $|x_i - x_i^*| \ge \sqrt{(4c\varepsilon/d)}$. Take $v = |x_i - x_i^*|$. Then, the above two inequalities allow us to conclude that

$$f(x_i^*, k_i) + \frac{dv^2}{2} < f(x_i, k_i) \text{ next, using } v \ge \sqrt{(4c\varepsilon/d)}$$
$$f(x_i^*, k_i) + 2c\varepsilon < f(x_i, k_i) \text{ which can be rearranged as}$$
$$f(x_i^*, k_i) + c\varepsilon < f(x_i, k_i) - c\varepsilon.$$

Using Lipschitz continuity for the LHS and RHS we get

$$f(x_i^*, k_i) + c\varepsilon > f(x_i^*, k_i^*)$$
$$f(x_i, k_i) - c\varepsilon < f(x_i, k_i^*).$$

Using the above two we get

$$f(x_i^*, k_i^*) < f(x_i, k_i^*).$$

By the definition of x_i^*, this is a contradiction. Therefore, $|x_i - x_i^*| < \sqrt{(4c\varepsilon/d)}$. □

If we assume both grading and doing homeworks have diminishing returns, it is very reasonable to say that (for a given $k_i > 0$) $H_i(x)$ and $(1 - k_i M_i(x))$ are both strictly monotonically increasing concave nonnegative functions. For shorthand, let $g(x) = (1 - k_i M_i(x)$.

Therefore, $H_i'(x) > 0$ and $H_i''(x) < 0$. Similarly, $g'(x) > 0$ and $g''(x) < 0$. The first derivative of $H_i(1-x)g(x)$ is $H_i(1-x)g'(x) - H_i'(1-x)g(x)$. The second derivative is equal to $H_i''(1-x)g(x) - 2H_i'(1-x)g'(x) + H_i(1-x)g''(x)$. We can see that it will always be negative.

We also assume the second derivative is below a certain value $-d$. If both of our functions are strictly increasing by at least a certain value d this assumption would always be true (because $-2H_i'(1-x)g'(x) < -2d^2$). The intuition behind this assumption is that for a fixed amount of work, the student will always reap a minimum amount of benefit.

Using this result, we characterize when the instructor's utility is close to optimal.

Theorem 4.6. *If for all i H_i and M_i are Lipschitz continuous with constant e and $|x_i - x_i^*| < \sqrt{(4c\varepsilon/d)}$, we have the following bound*

$$|U^{INST}(x_1, \dots x_n) - U^{INST}(x_1^*, \dots x_n^*)| < \sum_{i=1}^{n} (\alpha_i + \beta_i) e \sqrt{(4c\varepsilon/d)}$$

Proof. Recall that our instructor's utility is given by:

$$U^{INST}(x_1, \dots x_n) = \sum_{i=1}^{n} (\alpha_i H_i(x_i) - \beta_i M_i(1 - x_i)).$$

Note that if both H_i and M_i are Lipschitz continuous, then

$$|x_i - x_i^*| < \sqrt{(4c\varepsilon/d)}$$

implies that for some constant value e:

$$|H_(x_i) - H_i(x_i^*)| < e\sqrt{(4c\varepsilon/d)}$$
$$|M_(x_i) - M_i(x_i^*)| < e\sqrt{(4c\varepsilon/d)}$$

Therefore, we bound the instructor utility from each student

$$|(\alpha_i H_i(x_i) - \beta_i M_i(x_i)) - (\alpha_i H_i(x_i^*) - \beta_i M_i(x_i^*))| < (\alpha_i + \beta_i) e \sqrt{(4c\varepsilon/d)}$$

Thus, the overall utility is bounded as

$$|U^{INST}(x_1, \dots x_n) - U^{INST}(x_1^*, \dots x_n^*)| < \sum_{i=1}^{n} (\alpha_i + \beta_i) e \sqrt{(4c\varepsilon/d)}$$

□

The above guarantee is of the order $n\sqrt{\varepsilon}$. Given any constant ε' we can choose b such that $2^{-b} = \varepsilon$ is of order $(\varepsilon')^2/n^2$, giving approximation ε' and running in time $poly(n^5, (\varepsilon')^{-4})$.

5 Discussion

The largest barrier to implementing our model in real world scenarios is discovering the utilities of the students. Each student only provides a small number of observations to this function, since students are only in MOOCs for a limited amount of time, ruling out some learning techniques recently proposed [Blum *et al.*, 2014]. Each observation will likely be noisy, as there are many hidden variables that can affect how a student performs week to week. Each observation will likely be biased, as each week their homework and gradees change. Putting these observations together into a coherent utility function is a challenge. Previous work on MOOCs [Piech *et al.*, 2013] was able to get an approximate idea of how good each student was as a grader and how well each student performed, but there still was noise. Any system that attempts to learn the utilities would bound to have noise as well (such as [Nguyen *et al.*, 2014]), which we do not consider here.

Another barrier is communicating changes in audit level in a way that would be relevant to the students. Realistically, students are not going to differentiate between a .051 and a .052 chance of being audited. They might react if they are told that they are being "highly" vs "lowly" audited, or if their chance of being audited has increased. If we were dealing with automated systems, these small changes may matter.

In this model, we assume that students can only improve their grading at the cost of working on homework and can only improve their homework score at the cost of working on grading. There is no way the instructor can motivate a student to allocate more time in total. Expansions on our model might want to add a parameter that represents total time spent on the class instead of capping the total at a normalized value of 1.

Our work could be expanded to scenarios beyond peer grading. We could potentially consider any model where targets have to divide their time between a selfish action and a non-selfish action. For example, this could be extended to a factory dividing their time between making a product and cleaning up their industrial waste. All of our results would carry over, with possibly an easier way of determining utility functions.

6 Conclusion

We introduced a novel way of modelling how audits can incentivize desired student behavior in MOOCs. We argued that our model is realistic, and showed that computing an approximate equilibrium is computationally feasible.

Our work represents a new paradigm for strategic peer-grading. MOOCs require trustworthy grading and helpful feedback. By considering the strategic decisions each student has to make, course designers can better understand how their actions influence students' behavior and choose mechanisms to improve the experience of students within the limited TA resources.

7 Acknowledgements

This research was partially supported by the National Science Foundation (NSF) grants CCF0424422 and CNS1064688.

References

[An *et al.*, 2011] B. An, M. Tambe, F. Ordóñez, E.A. Shieh, and C. Kiekintveld. Refinement of strong stackelberg equilibria in security games. In *Proceedings of the 25th AAAI conference on Artificial Intelligence (AAAI'11)*, 2011.

[Birkhoff, 1946] Garrett Birkhoff. Three observations on linear algebra. *Univ. Nac. Tucumán. Revista A*, 5:147–151, 1946.

[Blocki *et al.*, 2013] J. Blocki, N. Christin, A. Datta, A.D. Procaccia, and A. Sinha. Audit games. In *Proceedings of the 23rd international joint conference on Artificial Intelligence (IJCAI'13)*, pages 41–47, 2013.

[Blocki *et al.*, 2015] J. Blocki, N. Christin, A. Datta, A.D. Procaccia, and A. Sinha. Audit games with multiple defender resources. In *Proceedings of the 29th AAAI conference on Artificial Intelligence (AAAI'15)*, 2015.

[Blum *et al.*, 2014] Avrim Blum, Nika Haghtalab, and Ariel D Procaccia. Learning optimal commitment to overcome insecurity. In *Advances in Neural Information Processing Systems*, pages 1826–1834, 2014.

[Garey and Johnson, 1979] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., 1979.

[Ghosh and Kleinberg, 2013] Arpita Ghosh and Jon Kleinberg. Incentivizing participation in online forums for education. In *Proceedings of the 14th ACM conference on Electronic commerce (EC'13)*, pages 525–542. ACM, 2013.

[Jain *et al.*, 2010] M. Jain, J. Tsai, J. Pita, C. Kiekintveld, S. Rathi, M. Tambe, and F. Ordóñez. Software assistants for randomized patrol planning for the LAX airport police and the federal air marshal service. *Interfaces*, 40(4):267–290, 2010.

[Jurca and Faltings, 2003] R. Jurca and B. Faltings. An incentive compatible reputation mechanism. In *Proceedings of the 5th IEEE International Conference on E-Commerce (CEC'03)*, pages 285–292. IEEE, 2003.

[Kiekintveld *et al.*, 2009] C. Kiekintveld, M. Jain, J. Tsai, J. Pita, F. Ordóñez, and M. Tambe. Computing optimal randomized resource allocations for massive security games. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'09)*, pages 689–696, 2009.

[Korzhyk *et al.*, 2011] D. Korzhyk, V. Conitzer, and R. Parr. Security games with multiple attacker resources. In *Proceedings of the 22nd international joint conference on Artificial Intelligence (IJCAI'11)*, pages 273–279, 2011.

[Li and Conitzer, 2013] Y. Li and V. Conitzer. Game-theoretic question selection for tests. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13)*, pages 254–262, 2013.

[Nguyen *et al.*, 2014] Thanh Hong Nguyen, Albert Xin Jiang, and Milind Tambe. Stop the compartmentalization: Unified robust algorithms for handling uncertainties in security games. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS '14, 2014.

[Piech *et al.*, 2013] C. Piech, J. Huang, Z. Chen, C. Do, A. Ng, and D. Koller. Tuned models of peer assessment in MOOCs. *arXiv preprint arXiv:1307.2579*, 2013.

[Shalem *et al.*, 2014] B. Shalem, Y. Bachrach, J. Guiver, and C.M. Bishop. Students, teachers, exams and MOOCs: Predicting and optimizing attainment in web-based education using a probabilistic graphical model. In *Proceedings of the 7th European Conference on Machine Learning and Knowledge Discovery in Databases (ECML/PKDD'14)*, pages 82–97, 2014.

[Tambe, 2011] M. Tambe. *Security and game theory: Algorithms, deployed systems, lessons learned*. Cambridge University Press, 2011.

[Vallam *et al.*, 2015] R.D. Vallam, P. Bhatt, D. Mandal, and Y. Narahari. A stackelberg game approach for incentivizing participation in online educational forums with heterogeneous student population. In *Proceedings of the 29th AAAI conference on Artificial Intelligence (AAAI'15)*, 2015.

[Wright *et al.*, 2015] J. Wright, C. Thornton, and K. Leyton-Brown. Mechanical ta: Partially automated high-stakes peer grading. In *Proceedings of the 32nd ACM Technical Symposium on Computer Science Education (ACM-SIGCSE'15)*, 2015.