Trust-aware Peer Assessment using Multi-armed Bandit Algorithms

Hou Pong Chan, Tong Zhao, and Irwin King
Department of Computer Science and Engineering
The Chinese University of Hong Kong, Hong Kong
{hpchan, tzhao, king}@cse.cuhk.edu.hk

ABSTRACT

Massive Open Online Coursers (MOOCs) offer a convenient way for people to access quality courses via the internet. However, the problem of grading open-ended assignments at such a large scale still remains challenging. Although peer assessment have been proposed to handle the large-scale grading problem in MOOCs, existing methods still suffer several limitations: (1) most current peer assessment research ignore the importance of how to allocate the assessment tasks among peers, (2) existing approaches for peer grading learn the complete ranking in an offline manner, (3) theoretical analysis for trust-aware peer grading is missing. In this work, we consider the case that we have prior knowledge about all students' reliability. We formulate the problem of peer assessment as a sequential noisy ranking aggregation problem. We derive a trust-aware allocation scheme for peer assessment to maximize the probability of constructing a correct ranking of assignments with a budget constraint. Moreover, we also derive an upper bound for the probability of prediction error on the inferred ranking of assignments. Furthermore, we propose the Trust-aware Ranking-based Multi-armed Bandit Algorithms to sequentially allocate the assessment tasks to the students based on the derived allocation scheme and learn an accurate peer grading result by taking students' reliability into consideration.

1. INTRODUCTION

Massive Open Online Courses (MOOCs) are indispensable as a means of online education by offering quality courses to students and sharing knowledge among people. With the great convenience of online education, the scales of MOOCs are always enormous and a MOOC from a famous university or a famous lecturer can easily attract more than tens of thousands of students to enroll. Hence, a challenging problem in MOOCs is the infeasibility of grading assignments or examinations using traditional TA/lecturer grading patterns. Due to this limitation, most MOOCs currently only offer assignments that can be graded automatically, such as multiple choice questions. However, the problem of grading open-ended assignments such as essay in MOOCs still remains challenging.

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2891080

To address this issue, peer assessment is proposed to handle the large-scale grading problem in MOOCs [13]. In peer assessment, each student grades a subset of his/her peers' assignments and submit their assessment results to the TA/lecturer. The final grading will be concluded based on the results of peer grading process.

Although peer grading shows a great efficacy in solving grading problems in MOOCs [16], there are still several issues in current peer grading methods.

- **First, most of current peer grading research focus on the aggregation of peer assessment results but ignore the importance of how to allocate the assessment tasks among peers.** Since different students have different capacities in grading assignments and they may also have different credibilities of their grading results, an arbitrary allocation of peer assignments may lead to uncertain and inaccurate peer grading results and also bring too much workload for some students.

- **Second, existing approaches for peer grading learn the complete ranking in an offline manner.** That is, the full ranking of assignments is learned after *all* the partial rankings have been collected [18]. Due to the offline manner of learning, one cannot take advantage of the sequentially gathered feedback to reduce the sample complexity. As a result, how to aggregate these partial orders sequentially and efficiently with a theoretical guarantee is a challenging and significant task.

- **Third, a solid theoretical analysis for trust-aware peer grading is missing.** To obtain an accurate complete ranking from peer grading, we need to estimate the reliability of each student. Unreliable students may give lower ranks to good assignments in order to increase their own relative scores in the course. Such malicious behaviors will lead to an inaccurate inferred ranking. There are some existing works [14, 12, 13, 15] incorporates grader reliability into their peer assessment models. Although they provide empirical evaluations, they do not give any theoretical bound on the sample complexities and the confidence value.

In this work, we investigate the problem of assessment tasks allocation of peer assessment. In our model, we consider the case that we have prior knowledge about all students' reliability from a separate reliability evaluation component. Our goal is to reduce the assessment workload of students (minimize the total number of peer assessments, known as sample complexity) by using a carefully designed allocation scheme and achieve an accurate peer grading result by taking students' reliability into consideration. We formulate the problem of peer assessment as a sequential noisy ranking

aggregation problem. Then, we derive a trust-aware strategy for peer assessment allocation with a budget constraint, the assessment workload of each student will not exceed a certain value. After that, we provide a theoretical upper bound of probability of prediction error on the inferred ranking of assignments. Finally, we propose the Trust-aware Ranking-based Multi-armed Bandit Algorithm to allocate the peer assessment tasks to the students based on the derived allocation scheme and aggregate the full ranking using a merge-sort based approach.

2. RELATED WORK

Existing peer assessment approaches can be divided into two main groups: *cardinal peer assessment* and *ordinal peer assessment*.

In cardinal peer assessment, each student gives cardinal grades, e.g. B+, to their peers' assignments. Traditional methods for peer assessments used naive methods such as taking median or mean of the received cardinal scores. [13] proposed a probabilistic learning algorithm to improve the accuracy of peer grading. However, since students are not well trained for grading assignments, different students may have different grading standards. For example, a student thinks that an assignment is deserved an A- grade, whereas another student thinks that the same assignment is only worth a B grade.

Ordinal peer assessment has been proposed to address the problem of diversified grading standards of students [16]. Instead of giving cardinal grades, each student ranks a subset of his/her peers' assignments, e.g. $a \succ c \succ e$. All the partial rankings are then aggregated to compute a complete ranking of all the assignments. Classical probabilistic ranking models such as the Plackett-Luce model [10], the Bradley & Terry model [2], and the Mallows model [11], were used to learn the full ranking of all the assignments based on the partial rankings of assignments collected from students[16, 14].

To obtain an accurate result from peer grading, we also need to estimate the reliability of students. Several existing methods introduced a variability parameter into their probabilistic models to estimate the grader reliability [13, 14, 12]. Although they provided an empirical evaluation using a peer grading dataset collected from a real class, they did not give any theoretical bound on the sample complexities and the confidence value.

In the theoretical aspect, ordinal peer grading can be viewed as a noisy ranking aggregation problem. Recent ranking aggregations studies transform the offline learning manner to an online and sequential manner by formulating the task as a Dueling Bandit problem [3, 20, 19, 17], which is a variant of Multi-armed Bandit (MAB) problem.

The "multi-armed bandits" problem refers to the problem a gambler faces at a row of slot machines, or "one-armed bandits", that look identical at first, but produce different expected rewards [1]. The crucial issue is to trade off acquiring new information (exploration) and capitalizing on the information available so far (exploitation). Dueling Bandit is a variant of Multi-armed Bandit problem, instead of directly observing the reward of one arm, we can only observe the result of a pairwise comparison between the rewards of two arms [20].

However, we cannot directly apply these existing dueling bandit approaches to peer grading because of the following limitations: (1) The existing Dueling Bandit setups do not consider the reliability of the grader and in fact, most of existing dueling bandits assume that there is only one grader, (2) The existing sequential setups do not consider the equality of two items, and (3) No existing method explores how to sequentially aggregate a set of partial orders consisting of more than two items for each.

3. PRELIMINARIES

We assume that there exists a true ranking of all the assignments. Suppose there are M students, let $a_1, ..., a_M$ be a set of assignments to be ranked. A ranking is represented as a bijection $r : \{a_1, ..., a_M\} \rightarrow \{1, ..., M\}$, which maps each assignment to its rank. Thus, $r(a_i)$ is the rank of a_i and $r^{-1}(i)$ is the item with rank i. For simplicity, we define $r_i = r(a_i)$ and $r_i^{-1} = r^{-1}(i)$. Let $\mathbb{S}_M$ denotes the set of all possible rankings of M assignments. We assume the occurrence of a particular ranking r follows a probabilistic distribution $\mathbb{P} : \mathbb{S}_\mathbb{M} \rightarrow [0, 1]$, then the probability that assignment a_i is better than a_j is defined as following.

$$\mu_{ij} = \mathbb{P}(a_i \succ a_j) = \sum_{r \in \{\mathbb{S}_M \mid a_i \succ a_j\}} \mathbb{P}(r) \tag{1}$$

If $\mu_{ij} > \frac{1}{2}$, then $a_i \succ a_j$ in the true ranking, i.e. a_i is better than a_j. If $\mu_{ij} < \frac{1}{2}$, then $a_j \succ a_i$ in the true ranking.

Our goal is to find the most probable ranking of all the assignments, which is defined as following.

$$r^* = \arg\max_{r \in \mathbb{S}_M} \mathbb{P}(r) \tag{2}$$

In the concrete implementation, we assume the probability distribution $\mathbb{P}$ on $\mathbb{S}_M$ follows Mallows model [11], which is a well known probabilistic ranking model. As shown in Equation 3, Mallows Model is parameterized by two parameters: dispersion ϕ and reference ranking $\tilde{r}$.

$$\mathbb{P}(r|\phi, \tilde{r}) = \frac{1}{Z(\phi)} \phi^{d(r,\tilde{r})} \tag{3}$$

The parameter $\phi \in (0, 1]$ controls the dispersion of the probability distribution. If $\phi = 1$, the probability distribution of rankings is uniform. If $\phi \rightarrow 0$, the probability distribution of rankings concentrates around the reference ranking, i.e. $\mathbb{P}(\tilde{r}|\phi, \tilde{r}) \rightarrow 1$. Thus, the reference ranking $\tilde{r}$ is the mode ranking of $\mathbb{P}$, we consider it as the true ranking of all the assignments.

$Z(\phi)$ is a normalization factor which ensures that the sum of the probability of all rankings equals to one. As shown in Equation 4, it only depends on the dispersion parameter ϕ [5].

$$Z(\phi) = \sum_{r \in \mathbb{S}_M} \mathbb{P}(r|\phi, \tilde{r}) = \prod_{i=1}^{M-1} \sum_{j=0}^{i} \phi^j \tag{4}$$

The distance function $d(r, \tilde{r})$ is the Kendall tau rank distance [8], which measures the dissimilarity between two rankings [3]. The Kendall tau distance is defined as the following, where r_i is the rank of assignment i in a particular ranking r.

$$d(r, \tilde{r}) = \sum_{1 \leq i < j \leq M} \mathbb{I}\{(r_i - r_j)(\tilde{r}_i - \tilde{r}_j) < 0\} \tag{5}$$

4. TRUST-AWARE ONLINE RANKING ELICITATION

4.1 Trust-aware MAB framework

With the aforementioned probabilistic assumptions, we can formulate the Trust-aware Multi-armed Bandit (MAB) framework as follows. In each iteration $t = 1, 2, ..., T$

1. The algorithm $\mathcal{A}$ allocates two students, i and j's assignments (a_i, a_j) to a student k to compare.

2. $\mathcal{A}$ observes a noisy binary feedback $o_{ijk} \in \{0, 1\}$, indicating which assignment is better than the other, e.g. $o_{ijk} = 1$ means $a_i \succ a_j$.

3. Based on the new observation o_{ijk}, $\mathcal{A}$ updates the value of $\hat{\mu}_{ij}^t$, which is the empirical proportion of "wins" of a_i against a_j by time t.

4. $\mathcal{A}$ determines the next assessment tasks allocation.

On the basis of this Multi-armed Bandit framework, if we want to infer the most probable ranking of all the assignments, the algorithm $\mathcal{A}$ would need to solve a noisy sorting problem. Hence, we devise an algorithm based on the well known merge sort algorithm and we will discuss our proposed algorithm in details in section 4.3.

4.2 Trust-aware allocation scheme

In our model, we assume that we have prior knowledge about all students' reliability from a separate reliability evaluation component. The reliability evaluation component estimates the trust value of each student, denoted by c_k for all students $k \in [1, M]$. Each trust value is fixed in the interval $(0, 1]$, which represents a belief about the probability that a student k gives a correct judgement.

Then, we estimate the trust-aware empirical probability that $a_i \succ a_j$ in the reference ranking using the trust value of each student as the corresponding weight. Equation 6 shows the estimation of the trust-aware empirical probability that $a_i \succ a_j$ at iteration t, where n_{ijk}^t is the number of comparisons between a_i and a_j judged by a student k among the first t iterations. Equation 7 shows the estimation of the trust-aware empirical probability that $a_i \succ a_j$ at the end of the algorithm, where n_{ijk} is the number of comparisons between a_i and a_j judged by a student k at end of the algorithm.

$$\hat{\mu}_{ij}^t = \frac{\sum_{k=1}^{M} n_{ijk}^t c_k o_{ijk}}{\sum_{k=1}^{M} n_{ijk}^t} \tag{6}$$

$$\hat{\mu}_{ij} = \frac{\sum_{k=1}^{M} n_{ijk} c_k o_{ijk}}{\sum_{k=1}^{M} n_{ijk}} \tag{7}$$

As we will use a merge-sort based algorithm to compute the most probable ranking of all the assignments, we need to decide the order of every pair of assignments to be compared with high probability. Hence, we need to find a confidence interval for our estimation. By using Hoeffding Inequality [6], we can obtain the confidence interval ϵ_{ij} for the trust-aware empirical probability $\hat{\mu}_{ij}$. Equation 8 shows the upper bound of the probability that our estimated answer $\hat{\mu}_{ij}$ deviates from the expected answer μ_{ij} by ϵ_{ij}, where μ_{ij} is the expected probability that $a_i \succ a_j$ in the reference ranking.

$$\Pr(|\hat{\mu}_{ij} - \mu_{ij}| \geq \epsilon_{ij}) \leq 2\exp(-\frac{2\epsilon_{ij}^2(\sum_{k=1}^{M} n_{ijk})^2}{\sum_{k=1}^{M} n_{ijk} c_k^2}) \tag{8}$$

We let $2\exp(-\frac{2\epsilon_{ij}^2(\sum_{k=1}^{M} n_{ijk})^2}{\sum_{k=1}^{M} n_{ijk} c_k^2} = \delta$, where $\delta \in [0, 1]$, such that with probability at least $1 - \delta$, it holds for particular a_i and a_j, $\widehat{\mu}_{ij} \in [\mu_{ij} - \epsilon_{ij}, \mu_{ij} + \epsilon_{ij}]$. Thus, we express the confidence interval ϵ_{ij} as follows,

$$\epsilon_{ij} = \sqrt{\frac{\log \frac{2}{\delta} \sum_{k=1}^{M} n_{ijk} c_k^2}{2(\sum_{k=1}^{M} n_{ijk})^2}} \tag{9}$$

Moreover, if we set ϵ_{ij} as follows,

$$\epsilon_{ij} = \sqrt{\frac{\log \frac{4MC_M}{\delta} \sum_{k=1}^{M} n_{ijk} c_k^2}{2(\sum_{k=1}^{M} n_{ijk})^2}} \tag{10}$$

We can obtain a restrict condition that for any pair of a_i and a_j, $\widehat{\mu}_{ij} \in [\mu_{ij} - \epsilon_{ij}, \mu_{ij} + \epsilon_{ij}]$ with probability at least $1 - \frac{\delta}{C_M}$, where $C_M = \lceil M \log_2 M - 0.91392M + 1 \rceil$ is the upper bound of the number of comparisons of the two-way top-down merge sort algorithm in the worst case performance [4, Theorem 1].

Based on Equation 7 and 10, our goal is to bound the total number of comparisons, n_{ijk}, while minimize the probability of prediction error on the order of every pair of assignments to be compared. Therefore, we need to express the probability of prediction error in terms of n_{ijk}. Without loss of generality, we consider the case when a_j is better than a_i in the reference ranking. Then, according to [3, Corollary 3], we have $\mu_{ij} \leq \frac{\phi}{1+\phi} < \frac{1}{2}$. Suppose we have a prediction error on the order of a_i and a_j, i.e. $(\hat{r}_i - \hat{r}_j)(\tilde{r}_i - \tilde{r}_j) < 0$, where $\hat{r}_i$ and $\hat{r}_j$ are the predicted rankings of a_i and a_j respectively. Then, $\widehat{\mu}_{ij} \geq \frac{1}{2}$ and $\widehat{\mu}_{ij} \in [\mu_{ij} - \epsilon_{ij}, \mu_{ij} + \epsilon_{ij}]$, we express the error probability $\Pr((\hat{r}_i - \hat{r}_j)(\tilde{r}_i - \tilde{r}_j) < 0)$ as follows,

$$\begin{aligned}
&\Pr((\hat{r}_i - \hat{r}_j)(\tilde{r}_i - \tilde{r}_j) < 0) \\
=&\frac{\mu_{ij} + \epsilon_{ij} - \frac{1}{2}}{2\epsilon_{ij}} \\
\leq&\frac{\frac{\phi}{1+\phi} + \epsilon_{ij} - \frac{1}{2}}{2\epsilon_{ij}} \\
=&\frac{1}{2} - (\frac{1}{2} - \frac{\phi}{1+\phi})\frac{1}{\epsilon_{ij}} \\
=&\frac{1}{2} - (\frac{1-\phi}{2(1+\phi)})\frac{1}{\epsilon_{ij}} \\
=&\frac{1}{2} - \frac{1-\phi}{2(1+\phi)}\sqrt{\frac{2(\sum_{k=1}^{M} n_{ijk})^2}{\log(\frac{4MC_M}{\delta})\sum_{k=1}^{M} n_{ijk} C_k^2}}
\end{aligned} \tag{11}$$

Next, we formulate an objective function to minimize $\Pr((\hat{r}_i - \hat{r}_j)(\tilde{r}_i - \tilde{r}_j) < 0)$ on every assignment pairs (a_i, a_j) to be compared as follows,

$$\begin{aligned}
&\underset{n_{ijk}}{\text{minimize}} \sum_{i,j} \Pr((\hat{r}_i - \hat{r}_j)(\tilde{r}_i - \tilde{r}_j) < 0) \\
\iff &\underset{n_{ijk}}{\text{minimize}} -\sum_{i,j} \frac{1-\phi}{2(1+\phi)}\sqrt{\frac{2(\sum_{k=1}^{M} n_{ijk})^2}{\log(\frac{4MC_M}{\delta})\sum_{k=1}^{M} n_{ijk} C_k^2}} \\
\iff &\underset{n_{ijk}}{\text{minimize}} -\sum_{i,j} \frac{1-\phi}{2(1+\phi)}\sqrt{\frac{2}{\log(\frac{4MC_M}{\delta})}}\sqrt{\frac{(\sum_{k=1}^{M} n_{ijk})^2}{\sum_{k=1}^{M} n_{ijk} C_k^2}}
\end{aligned} \tag{12}$$

This objective function is not necessary convex. Since $c_k \in (0, 1]$, $\sum_{k=1}^{M} n_{ijk} \geq \sum_{k=1}^{M} n_{ijk} c_k^2$. We can relax $-\sqrt{\frac{(\sum_{k=1}^{M} n_{ijk})^2}{\sum_{k=1}^{M} n_{ijk} C_k^2}}$ to its upperbound $-\sqrt{\sum_{k=1}^{M} n_{ijk} C_k^2}$ [9]. This relaxation results in a convex objective function as follows,

$$\iff \underset{n_{ijk}}{\text{minimize}} -\sum_{i,j} \frac{1-\phi}{2(1+\phi)}\sqrt{\frac{2}{\log(\frac{4MC_M}{\delta})}}\sqrt{\sum_{k=1}^{M} n_{ijk} C_k^2} \tag{13}$$

After ignore all those constant terms, we formulate the objective function as shown in Equation 14. Moreover, we also impose several constraints on it. First of all, n_{ijk} cannot be negative, as shown in Equation 15. Moreover, we would like a more reliable student to grade more assignments, as shown in Equation 16. [13] conducts an experiment to model the relations between the students' scores and their reliability, and the results showed that students with higher

scores tend to have higher reliability. On the basis of the result, we assume that a student with a higher reliability tends to have a higher score. Hence, if the difference of reliability between two students i and j is large, we can assume that the difference of their scores is also high, so the number of pairwise comparisons between their assignments, i.e. $\sum_{k=1}^{M} n_{ijk}$ can be small. Moreover, the total number of comparisons cannot exceeds a peer assessment workload limit B, as shown in Equation 17. The complete optimization problem is shown as follows.

$$\underset{n_{ijk}}{\text{minimize}} \quad -\sum_{i,j}\sqrt{\sum_{k} n_{ijk}c_k^2} \tag{14}$$

$$\text{subject to} \quad n_{ijk} \geq 0,\ \forall i,j,k,\ i \neq j \neq k. \tag{15}$$

$$(\sum_{i,j} n_{ijk} - \sum_{i,j} n_{ijk'})(c_k - c_{k'}) \geq 0$$

$$\forall i,j,k,k' \in [1,M],\ i \neq j \neq k \neq k' \tag{16}$$

$$\sum_{ij}[(\sum_{k} \frac{n_{ijk}}{c_k})(c_i - c_j)^2] \leq B$$

$$\forall i,j,k \in [1,M]\ i \neq j \neq k \tag{17}$$

The analytic solution n^*_{ijk} of the above optimization problem is the required number of comparisons of a_i and a_j by student k that maximize the probability of constructing a full ranking correctly. As it is a convex optimization problem, we can use Karush-Kuhn-Tucker (KKT) conditions to solve the analytic solution [7]. Then, we derive the following allocation scheme,

THEOREM 1. *Assume the probability distribution of $\mathbb{P}$ on $\mathbb{S}_M$ follows Mallows model, and we have prior knowledge about all students' trust and these trust values are fixed in a known interval $\in (0,1]$. To maximize the probability of constructing a correct ranking of assignment with a budget constraint B, the required number of comparisons between a_i and a_j by student k is*

$$n^*_{ijk} = \begin{cases} \frac{(c_{i^*_k} - c_{j^*_k})^2 B}{\sum_{k'} \frac{(c_{i^*_{k'}} - c_{j^*_{k'}})^2}{c_{k'}}} & \text{if } i = i^*_k, j = j^*_k \\ 0 & \text{otherwise} \end{cases}$$

where $i^*_k, j^*_k = \arg\max_{i,j} \frac{(c_i - c_j)^2}{c_k}$ *and* $i,j,k = 1,2,...,M$

By substituting the n^*_{ijk} into Equation 11, we derive the following corollary,

COROLLARY 2. *The upper bound of the probability of prediction error* $\Pr((\hat{r}_i - \hat{r}_j)(\tilde{r}_i - \tilde{r}_j) < 0)$ *on every assignment pairs (a_i, a_j) to be compared in a two-way top-down merge sort algorithm is*

$$\frac{C_M}{2} - \sum_{i,j} \frac{1-\phi}{2(1+\phi)} \sqrt{\frac{2}{\log(\frac{4MC_M}{\delta})}} \sqrt{\sum_{k=1}^{M} \frac{(c_{i^*_k} - c_{j^*_k} + \alpha)^2 c_k^2 B}{\sum_{k'} \frac{(c_{i^*_{k'}} - c_{j^*_{k'}})^2}{c_{k'}}}}$$

4.3 Trust-aware ranking based MAB Algorithm

With the required number of comparisons of a_i and a_j by student k, we can devise an efficient algorithm to sequentially allocate the assessment tasks with a trust-aware strategy and learn an accurate ranking of all the assignments by taking students' reliability into consideration.

As shown in Algorithms 1, 2 and 3, the proposed algorithms are based on the two-way top-down merge sort algorithm and a Dueling Bandit algorithm, MALLOWSMPR[3]. Firstly, the TRUSTAWARERANKINGBASEDMAB procedure calls the SPLITMERGE procedure. Then, the procedure SPLITMERGE recursively splits the unordered set of assignments into smaller subsets until the size of the subset is 1. Then all the subsets are merged to a sorted list by calling the procedure TRUSTAWAREMERGE.

In the setting of a merge sort algorithm, we can directly observe the pairwise relation between two items. While in the scenario of peer assessment, we can only observe the pairwise relation between a_i and a_j by the peer assessment feedback provided by students. As a result, whenever the algorithm needs to know the pairwise relation between a_i and a_j, it allocates the assessment tasks of (a_i, a_j) to a student k according to the value of n^*_{ijk}.

However, the value of n^*_{ijk} may be larger than 1, and it is pointless to ask the same person to judge the same pair of assignments for multiple times. Therefore, we only choose student k to compare a pair of assignments (a_i, a_j) once if $n_{ijk} > 0$.

Algorithm 1 TRUSTAWARERANKINGBASEDMAB

```
for i = 1 → M do r_i = i, r'_i = 0
(r', r)=SPLITMERGE(r, r', 1, M)
for i = 1 → M do r'_i = i
return r
```

Algorithm 2 SPLITMERGE$(r, r', begin, end)$

```
if end − begin > 1 then
    mid = ⌊(begin + end)/2⌋
    (r, r') = SPLITMERGE(r, r', begin, mid)
    (r, r') =SPLITMERGE(r, r', mid, end)
    (r, r') = TRUSTAWAREMERGE(r, r', begin, mid, end)
    for l = begin → end do r_l = r'_l
return (r, r')
```

Algorithm 3 TRUSTAWAREMERGE$(r, r', begin, mid, end)$

```
l = begin, l' = mid
for q = begin → end do
    if (l < mid)&(l' ≤ end) then
        for k' = 1 → M do
            if n*_{ll'k'} > 0 then
                Allocate (a_l, a_{l'}) to student k'
                Observe o^t_{ll'k'} = I_{k'}{r_l < r_{l'}}
                μ̂_{l,l'} = μ̂_{l,l'} + o^t_{ll'k'}
        if 1/2 < μ̂_{l,l'} − ε_{l,l'} then
            r'_q = r_l, l = l + 1
        else
            r'_q = r_{l'}, l' = l' + 1
    else
        if (l < mid) then
            r'_q = r_l, l = l + 1
        else
            r'_q = r_{l'}, l' = l' + 1
return (r, r')
```

5. CONCLUSION AND FUTURE WORK

In this work, we study the problem of trust-aware peer assessment in MOOCs. To address the limitations of existing peer assess-

ment methods, we transform the current offline learning manner of peer assessment into an online and sequential manner.

We derive a trust-aware allocation scheme to allocate peer assessment tasks to students while maximizing the probability of constructing a correct ranking of assignments with a budget constraint. Moreover, we also derive the upper bound of the corresponding probability of prediction error on the inferred ranking of assignments. Furthermore, we propose the Trust-aware Ranking-based Multi-armed Bandit Algorithms to sequentially allocate the assessment tasks to the students based on the derived allocation scheme and aggregate the full ranking using a merge-sort based approach.

However, the allocation scheme exhibits sparsity feature, in other words, some pairs of assignments may have no comparisons at all, while other pairs of assignments may have many comparisons. Hence, we will address this problem by introducing a regularization term in the objective function to penalize the sparse behavior and derive a new allocation scheme based on the new optimal solution. In the future, we will conduct experiments to evaluate the accuracy and efficiency of the proposed algorithms using both synthetic data and dataset in real-world MOOCs. Furthermore, we would like to collaborate with MOOCs providers and apply the proposed peer assessment framework into their MOOCs platforms. As the derived allocation scheme and the proposed algorithms assume that when we have prior knowledge about all students' reliability and the trust values are fixed in a known interval, we will consider the case that we do not have prior knowledge about all students' reliability. We will extend the proposed Trust-aware Ranking-based Multi-armed Bandit algorithm to learn reliability of all the students adaptively during the peer assessment process.

6. REFERENCES

[1] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *In Machine Learning*, 2002.

[2] R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs the method of paired comparisons. *Biometrika*, 39(3-4):324–345, 1952.

[3] R. Busa-Fekete, E. Hüllermeier, and B. Szörényi. Preference-based rank elicitation using statistical models: The case of mallows. *In ICML*, 2014.

[4] P. Flajolet and M. J. Golin. Mellin transforms and asymptotics: The mergesort recurrence. *Acta Inf.*, 31(7):673–696, 1994.

[5] M. A. Fligner and J. S. Verducci. Distance based ranking models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 359–369, 1986.

[6] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 1963.

[7] W. Karush. *Minima of functions of several variables with inequalities as side constraints*. PhD thesis, Master's thesis, Dept. of Mathematics, Univ. of Chicago, 1939.

[8] M. G. Kendall. Rank correlation methods. 1948.

[9] X. Liu, H. He, and J. S. Baras. Trust-aware optimal crowdsourcing with budget constraint. In *2015 IEEE International Conference on Communications, ICC 2015, London, United Kingdom, June 8-12, 2015*, pages 1176–1181, 2015.

[10] R. D. Luce. *Individual choice behavior: A theoretical analysis*. Courier Corporation, 2005.

[11] C. L. Mallows. Non-null ranking models. i. *Biometrika*, pages 114–130, 1957.

[12] F. Mi and D. Yeung. Probabilistic graphical models for boosting cardinal and ordinal peergrading in moocs. *In AAAI*, 2015.

[13] C. Piech, J. Huang, Z. Chen, C. B. Do, A. Y. Ng, and D. Koller. Tuned models of peer assessment in moocs. *In CoRR*, 2013.

[14] K. Raman and T. Joachims. Methods for ordinal peer grading. *In SIGKDD*, 2014.

[15] K. Raman and T. Joachims. Bayesian ordinal peer grading. *In L@S*, 2015.

[16] N. B. Shah, J. K. Bradley, A. Parekh, M. Wainwright, and K. Ramchandran. A case for ordinal peer-evaluation in moocs. *In NIPS Workshop on Data Driven Education*, 2013.

[17] B. Szörényi, R. Busa-Fekete, A. Paul, and E. Hüllermeier. Online rank elicitation for plackett-luce: A dueling bandits approach. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, R. Garnett, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 604–612. Curran Associates, Inc., 2015.

[18] F. Wauthier, M. Jordan, and N. Jojic. Efficient ranking from pairwise comparisons. *In ICML*, 2013.

[19] Y. Yue, J. Broder, R. Kleinberg, and T. Joachims. The k-armed dueling bandits problem. *In J. Comput. Syst. Sci.*, 2012.

[20] Y. Yue and T. Joachims. Interactively optimizing information retrieval systems as a dueling bandits problem. *In ICML*, 2009.