

Choosing Peers, Improve the Quality of Peer Assessment

Jinshuang Li^a, Xiaohan Fu and Qiuwei Yang

School of Computer Science and Engineering, Northeastern University, China

^alijinshuang@mail.neu.edu.cn

Keywords: peer assessment; evaluation quality; online education; performance

Abstract. Peer assessment has got a lot of attention because it serves as a useful tool for scaling the grading of open-ended works. But it does not always produce accurate results because of two important reasons: students' ability and students' attitude. Based on students' peer grades and the evaluation attitude in the previous tasks, we proposed a new method to calculate students' evaluation ability value. Then we improved user allocation algorithm. In this way, we can protect learning enthusiasm of all students. Meanwhile, we calculate final score through choosing good students. So we can enhance the credibility of peer grade and show students comments of better quality.

Introduction

There has been tens thousands or even hundreds thousands of people studied in the online course simultaneously. Learning assessment especially the subjective or capability valuation become a big problem. Therefore, the researchers have to think about the logistical on MOOCs. Peer assessment has gained broad attention and become a standard setup for majority MOOCs platforms.

Although a great deal of research shows that peer assessment can evaluate students' grade accurately, its acceptable levels by participated students is a problem. The assessment quality depends upon the participants' ability and attitude. Apparently most student can accomplish task, so they have the ability to evaluate peer task. But there are a still a few students who study not very well, so they don't have such ability. It is a truth that there are great gaps between most students and teachers in knowledge and ability. So students, especially excellent students, doubt that if others have the ability to evaluate their task. On the other hand, some students evaluate their tasks carelessly. Students who read a nonsense comments may doubt their grades, feel being treated unfairly, further make negative influence on motivation and finally affect teaching effects.

This paper proposes a method which reject the students with poor ability and attitude according to their performance in the previous tasks. We selected those students with good ability and attitude to participate in new task evaluation. It greatly improves the quality of the peer assessment.

Participants in the Experiment

This experiment is mostly targeted at the students of the course named Introduction to Computer Science in our university. The course started in October 2015 has 32 experimental hours and lasts for 6 weeks. There are 2303 students all from grade one involved in the experiment, covering 39 majors. The distribution of participants is shown in Table 1.

institute	The number of students
School of Mechanic	643
School of science	537
School of Metallurgy	172
School of Foreign Languages	647
Others	227

Table 1. The experiment results of the second task.

In the course of *Introduction to Computer Science*, students are supposed to learn to use a number of general software. The works include writing and typesetting science paper, calculating and Handling data, editing photos and videos etc. The learning of each software requires a lot of time to practice and train, therefore, we decided to use SPOCs model improving experimental quality. We provide all necessary martial to the students, including the video recording, text explanation of operating etc. It evaluates the learning quality of students by our peer assessment system.

We chose three tasks of the course to do the experiment, they are calculating and handling data, photo editing as well as scientific and technological paper writing and typesetting. Before the three tasks, students have done a simple peer assessment task, so they have been familiar with the whole evaluating operation procedure and how to use the evaluation system.

Every work has been assigned to 10 reviewer. And each assignment has 7-10 grading rubrics. During the review, students will give a value based on the quality of each rubric by a 1+5 scale. The items have statement instructing student selection. The items are respectively correspond the values from 0 to 5. The score 0 means nothing have been done according to the rubric; the score 1 means the quality is pretty awful; the score 2 represents a bad work; the score 3 is for a common level; the score 4 deserves a good work and the score 5 is supposed to be given to a perfect works.

Estimate the Ability of Graders

Quality of peer grade has a great relationship with the ability of participants. Although the graders have different ability, they have often get considerable professional knowledge and technical training through courses. For this reason, most of the students have the ability to evaluate the quality of the assignment works. But any popular teacher will have some students who do not study hard, so they don't have enough ability to accomplish peer assessment. The student emotion must be effected if we let these students alone, which is a common phenomenon. Selecting these students who have the enough ability to conduct peer grading is necessary to improve the credibility of the evaluation grade.

Obviously, students' grades reflect their ability. These students who get a high score are more likely to perform better in the next task. Especially in continuous works (knowledge and skill used to accomplish the current work are closely related to previous works). Thus predicting students' ability of evaluating new works by their previous peer grade is convincing. Which is used in our experiment. And we normalize their grades to ensure the value is between 0-1.

Estimating the student attitude value

Although students have the ability to evaluate the work of other students, some students do not have a responsibility, he do not evaluate the work seriously. It will affect the evaluation effectiveness. Generally the students are satisfied with the grade of their works, but when they see comments, they will feel that they are being treated unfairly. This result in dissatisfaction with the peer assessment. The comments is more teaching effect than peer assessment. It an important part in the peer assessment. So it is necessary to show students the comments.

Because it is unavoidable that a few students evaluate peer task carelessly, so it is reasonable extracting these students who can carefully evaluate their peer tasks. It can achieve better evaluated result. How to evaluate the students' attitude? We mainly consider two aspects: the students' grading bias and the quality of the students' comments.

Because the work is mainly aimed at the software operation technical ability. The rubrics is clear, such as "including five picture above, and one picture is about our campus". There is no considerable controversy, so if a student's peer value is deviation from mean value obviously, then the student can be identified either intentionally biased or just casually. So we use the standard error of the evaluation rubrics calculating their grading bias.

In the last part evaluation procedure, students have to write the comments about the work. It should include following the suggested: the advantages, disadvantages and the suggestions to improve the work, etc. This section uses the free-form writing, minimum number of words are

required. The peer assessment application can use the text analysis technology to know which students do not write comments seriously, it can judge the attitude of the evaluation of students. In our experiment, we use two mechanisms to evaluate the quality of the comments: one simple way is to count the number of words, it less than the required words will be punished, if a comment word number lower 10 percent will deduct 10 percent evaluation attitude score, and so on. Another way is the re-use of public supervision, students who do not satisfied with the comments can report it to teacher. Teachers will deal with every report and it will influence reporter or reportee's attitude scores.

In the last part of a complete evaluation of the work, requiring students to write the reviews of works following the suggested format, including the advantages and disadvantages and the areas for further improvement, suggestions, etc., the evaluation of the students plays a very large role in improving students' improve. This section uses the free-form writing, minimum number of words required, we can use the text analysis to know which students do not write comments seriously, so as to determine the attitude of the evaluation of students. We use two mechanisms to evaluate the reviews of scoring: one simple way is to count the number of words, number of words less than the required number of words per 10%, we will deduct 10% of the evaluation score; another is the re-use of public supervision the way, in the grades pre-release stage, students can see the evaluation of their works, for reviews have bigger dispute or dissatisfied evaluation attitude can be complaints, Complaints on the evaluation of the attitude of the specialized teachers to deal with, the results will be directly reacted to the evaluation of the opinion scores.

The grading bias and the comments attitude score should be normalized. The score of students' evaluation attitude is the weighted sum of scoring deviation value and comments attitude value.

Calculating the student review ability

When student s finish a task, we compute his review ability value as follow:

$$a_s = k_1 d_s + k_2 \alpha_s \tag{1}$$

Where a_s is the ability value of student s and d_s is the evaluation attitude value of student s .

In our experiments, we believe that the students' ability and attitude are equally important in the evaluation process, so we set the weight $k_1 = k_2 = 0.5$.

Most of the teachers assign more than one task in a course, it is intuitively to choose more tasks to calculate the student evaluation ability value more effective. Let indicates the current estimate value S_i when student s finish k th task, then we compute the student evaluation value Q_{k+1} using the equation as follow.

$$Q_{k+1} = \sum_{i=1}^k \omega_1 S_i + \omega_2 Q_k, Q_1 = S_1, \omega_1 + \omega_2 = 1 \tag{2}$$

In our experiments, after the first task we compute each students' evaluation ability value. Then we used it to cut the second task. After the second task, we compute each students' evaluation ability value again. Then we used upper equation to calculate the students' new value. We compute using the same ω_1 and ω_2 . Then we used it to cut the third task.

According to above method, we get a ranking of all the students, which is necessary for the allocation algorithm.

Hierarchical Allocation algorithm

According to the ranking about students' evaluation ability values, we can cut all the students who participate in this task to many levels. During the allocation process, we can extract the students according to different levels, which can effectively avoid the possibility of some works reviewed by too many poor students. It can also exclude the students who have poor evaluation ability so that improving user's satisfaction of peer assessment.

In our experiment, we divide the students into two levels: senior group and junior group. Although we can only extract some better students to participate in the peer assessment, peer evaluation is not only a way of evaluation, but also a pedagogical. Students can learn a lot in the evaluation process. In order to achieve the purpose of the pedagogical, we assign the same number of tasks for each students. From the student's point of view, everyone review the works. They should be responsible for the final grade of others.

For that reason, we have to ensure that each assignment is reviewed by certain number of senior students and junior students. i.e., if the number of reviewer for each assignment is m , we select a certain percentage of the senior students as the reviewers. For example, let m be 10, and the proportion of senior level students is 60%, we should ensure that each assignment will reviewed by 6 senior students and 4 junior students.

Conclusion

Based on students' peer grades and the evaluation attitude in the previous peer assessment tasks, proposed a new method to calculate students' evaluation ability value. We use assessment grade to calculate users' ability and then use standard deviation of each evaluation rubrics and the quality of evaluation comments to grade students' evaluation attitude. Through the above two aspects results, we can calculate users' evaluation ability value, then divide students into different levels.

We improved user allocation algorithm to ensure each assignment work shall be allocated by the level proportion of students. In this way, we can protect learning enthusiasm of all students. Meanwhile, we calculate final score through choosing good students. So we can enhance the credibility of peer assessment grade and show students comments of better quality. Peer assessment is not only a substitute tool for teacher evaluation, but also play an important role on pedagogical.

Acknowledgements

This work was financially supported by the Liaoning Science and Technology Projects (2013217004-1).

References

- [1] Masayuki Ashikawa, Takahiro Kawamura, Akihiko Ohsuga. Deployment of Private Crowdsourcing System with Quality Control Methods. In Proceedings of the 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), p.9-16
- [2] András Benedek, György Molnár, and Zoltán Szűts. Practices of Crowdsourcing in relation to Big Data Analysis and Education Methods. In Proceedings of IEEE International Symposium on Intelligent Systems & Informatics, p.1814-1819
- [3] Yukino Baba, Hisashi Kashima. Statistical Quality Estimation for General Crowdsourcing Tasks. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'13), p.554-562
- [4] Stewen Dow, Anand Kulkarni, Scott Klemmer, Björn Hartmann. Shepherding the Crowd Yields Better Work. 2012. In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW'12), p.1013-1022
- [5] Andrew Cross, Mydhili Bayyapunedi, Dilip Ravindran, Edward Cutrell, and William Thies. VidWiki: enabling the crowd to improve the legibility of online educational videos. In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (CSCW'14), p.1167-1175