Towards Better Grading: Promoting Grading Fairness with Assessment Decision Tree Visualization

Yuqian Chai, Xinyu Qi, Ling Li, Mansurbek Kushnazarov, Cheuk-Wang Yau, Yifei Dong, Chi-Un Lei
The University of Hong Kong
yqchai@eee.hku.hk, andreaq@hku.hk, lingli1000@gmail.com, mansurbek@teli.hku.hk, mcwyau@hku.hk, chloedong@teli.hku.hk, culei@hku.hk

ABSTRACT: Assessment fairness has been a challenging issue in educational practice, particularly for large-scale courses in which multiple graders are involved. In order to help guarantee grading fairness, this paper proposes an approach—'assessment decision tree visualization'—that enables graders to identify error cases and adjust the grading if necessary. After implementing the approach into a university course with 136 students, results showed such visualizations could help teachers achieve fairer grading practices.

Keywords: Assessment Fairness, Decision Tree, Visualization

1 INTRODUCTION

Assessment fairness has remained challenging in educational research and practice for decades. Literature has suggested that grading transparency and use of rubrics could lead to fairer assessment. However, in a large sized class where multiple graders are involved, graders may have different interpretations of a same rubric, resulting in grading discrepancy. To ensure grading fairness, quality check is usually carried out, which is however a time-consuming and laborious task. In this study, we propose to use assessment decision trees based on students' interaction data and grades to replace the manual check and achieve grading fairness without excessive effort.

2 EXPERIMENT

There is an ongoing trend of utilizing online interaction data in teaching and learning practices. It could offer an overview of students' learning performances and habits without too much effort required. Subsequently, if we select students' interaction data as the input and grades as the prediction target, we can build the assessment decision tree. The decision tree algorithm is a classical machine learning algorithm that can be applied in both classification and regression issues. One of the major advantages of the decision tree is that it is easy to interpret and visualize. Graders can adjust their grades by checking and comparing their assessment tree with others'.

2.1 Method

To better illustrate the process, we have conducted an experiment in an undergraduate level course with an enrollment size of 136. In the course, students needed to complete a graded activity which required them to work in small groups and answer open-ended questions in an online shared document. The submissions are then graded by ten graders based on each individual student's

contributions to the collective response. In other words, there should be a high correlation between students' interactions and grades. During the activity, students can either edit the documents directly to answer questions or create comments to discuss with teammates. Both actions are considered as contributions. Subsequently, we collected number of participated questions (*#questions*), number of comments (*#comments*), and total number of contributed words (*total*) as input features. Since there are four grade scales for this activity (0, 5, 7, 10), the decision tree we built is a classification tree. The average sample size of each grader's assessment decision tree is 68.

2.2 Results

We built each grader's assessment decision tree and two of them are presented in Figure 1. It is clear that Grader 2 was stricter than Grader 1 as students on the right tree are less likely to receive the highest score indicated by the root node. Specifically, Grader 1's decision tree indicates multiple scenarios in which students could receive the highest score (grade_10), while Grader 2's decision tree only suggests one of such scenarios with a small sample size of 3. The decision tree visualization was presented to the graders and most of them have modified their grades accordingly.

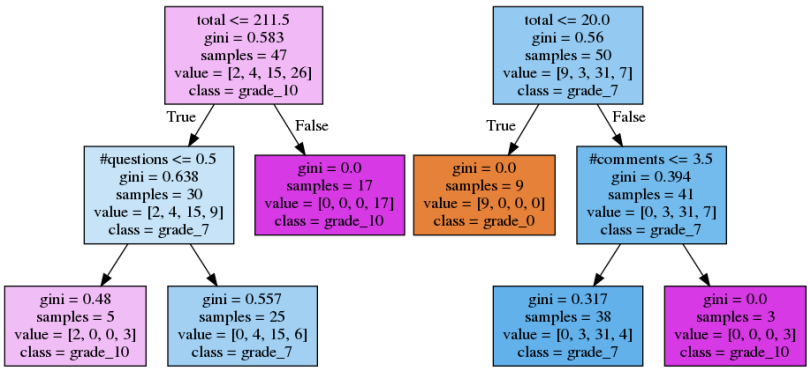

Figure 1: The assessment decision trees of Grader 1 (left) and Grader 2 (right)

3 LIMITATIONS

Despite the advantages of assessment decision trees, there are several limitations while applying this method. First of all, it is particularly applied in the evaluation of text-input participation. For other types of tasks, students' interaction data may not have a tied relationship with grades. In these cases, teachers can manually label some cognitive features, for example, frequency of grammar mistakes (seldom, sometimes, frequently). By doing so, the assessment decision tree can help verify the fairness in grading. Additionally, there is a sample size requirement for the decision trees, but it would not be an issue for large-scale courses.

REFERENCES

Tierney, R. D. (2013). Fairness in classroom assessment. *SAGE handbook of research on classroom assessment*, 125-144.

Magerman, D. M. (1995, June). Statistical decision-tree models for parsing. In *Proceedings of the 33rd annual meeting on Association for Computational Linguistics* (pp. 276-283). Association for Computational Linguistics.