Which Questions Have High Quality: Question Quality Evaluation Metric Based on Answer Records

Shiwei Tong[1], Ye Huang[1], Rui Lv[1], Wei Tong[2], Qi Liu[1],
Zhenya Huang[1], Yu Su[3], and Enhong Chen[1]

[1]Anhui Province Key Laboratory of Big Data Analysis and Application, School of Computer Science and Technology & School of Data Science,University of Science and Technology of China
{tongsw, huangyehy, lvrui2018}@mail.ustc.edu.cn; {qiliuql, huangzhy, cheneh}@ustc.edu.cn
[2]National Education Examination Authority
tongw@mail.neea.edu.cn
[3]iFLYTEK Research, iFLYTEK CO., LTD.
yusu@iflytek.com

Abstract

This paper presents our solution to the task 3 of the NeurIPS 2020 Education Challenge, which aims to predict the quality of questions. We propose a statistical evaluation metric to evaluate the quality of the questions. The evaluation metric mainly includes the difficulty and the correct rate of the question. A high-quality question should have moderate difficulty and correct rate, so the value is closer to 0.5, the higher the quality of the test. We use the Logistic model in item response theory to estimate the difficulty of the questions. The Logistic model considers the difficulty, the discrimination of questions, and the ability of students, but it has a deviation in the questions with fewer answer records. For mitigating this deviation, we count the correct rate of the questions and integrate the ranking results of these metrics as the final submission. Our proposed method has achieved the accuracy of 0.9600 on public evaluation and 0.7200 on private evaluation, which are the top1 and top3 performances respectively in this task.

Preprint. Under review.