

Computerized Adaptive Assessment

¹Dinesh B Satre, ²Adarsh Saji, ³Aniket Chavan, ⁴Aniket Dandade, ⁴Deepanjaya Modak

¹Assistant Professor, ²Student, ³Student, ⁴Student, ⁵Student

^{1,2,3,4,5}Computer Department,

^{1,2,3,4,5}Mrathwada Mitramandal's Institute of Technology, Pune, India.

Abstract : The estimation of the difficulty level of a given question is required in order to understand how well the student understood a particular topic, which would help the teachers to grade a student accordingly and decide how much attention is to be given to a particular student. There is no doubt that teachers often misjudge the difficulty level of the questions. In order to avoid that, a module is created in a data processing system for rating difficulty of a question. In this mechanism, a set of question bank is set by the teacher/admin whose difficulty level is judged by the basically two factors, first, the number of times the question is answered correctly by a particular number of students and second, the amount of time required in by the student to provide a sound analysis. The mechanism that generates the difficulty score of a question based on such above-mentioned factors uses a machine learning module. This machine learning module map features to assigned weights for scaling the difficulty score. This provided analysis could be very valuable for both teachers and students in order to detect unclear problem statements and rectify students' misconceptions.

IndexTerms - Machine Learning, Ordinal Logistic Regression, classify.

I. INTRODUCTION

In typical computerized assessments, students the test taker attempts random questions pre-assigned by the test administrator, here the issues are different among the range of test-takers, probably the test might have multiple sections, like easy, moderate, and difficult. The test-taker with deep knowledge might not find the test to be challenging enough, the test taker with moderate knowledge will find the test a little difficult, and the test taker with little knowledge will find the test very challenging.[1]

So to acknowledge this issue a computerized adaptive assessment is being designed which would adapt according to the level of the test taker's understanding. This module aims a fair chance for every type of test-takers, i.e the one with maximum knowledge to score the most by solving almost all challenging questions and the one ones with less knowledge to solve easy and moderate questions and get marks according to their knowledge[5][6]

To make the examination platform adaptive, a machine learning algorithm named ordinal logistic algorithm is applied, which would classify the questions according to the difficulty levels of the questions considering the number of students answering the question correctly and the time required to solve a particular question.

When the candidate solves a question correctly then the assessment system will assign a question of higher difficulty level to the test taker, if the current difficulty level is not the highest difficulty level. Similarly, if the test taker had solved a question incorrectly, then the current difficulty level assigned to the test taker is decreased by one unless the current level is not easy.

II. BACKGROUND

Right now, in practice Graduate Record Examination(GRE)[8], which is a standardized test which is a requirement for admission in various graduate schools in the United States and Canada, uses an adaptive assessment pattern, but it does not work exactly as the system explained in this paper. In GRE examination is conducted in two sets, where the set one and set two are of the same pattern the only difference is varying difficulty level. If the candidate solved the first set very well then the next set will be more difficult, or else the next set will be an easier one. Here the assignment of the difficulty level of the questions is done by the examiners/administrators which would not be accurate as per the students. Here in this paper, the difficulty level is set by the system according to the response of the previous test-takers. Also here the system explained in this paper, the allocation of the questions according to the previous question is done, it does not wait for an entire set to complete.

III. WORKING OF THE SYSTEM

a. Question Difficulty Calculation

The candidates attempt the questions parallelly. After each candidate submits an answer to a question, the difficulty level assigning module looks into the current difficulty factor of the question and the time required to solve the question, then according to the response, the difficulty level is updated after performing the required calculations. Here the faculty/admin approach module is there to assign the difficulty level of a totally new question, which is new in the

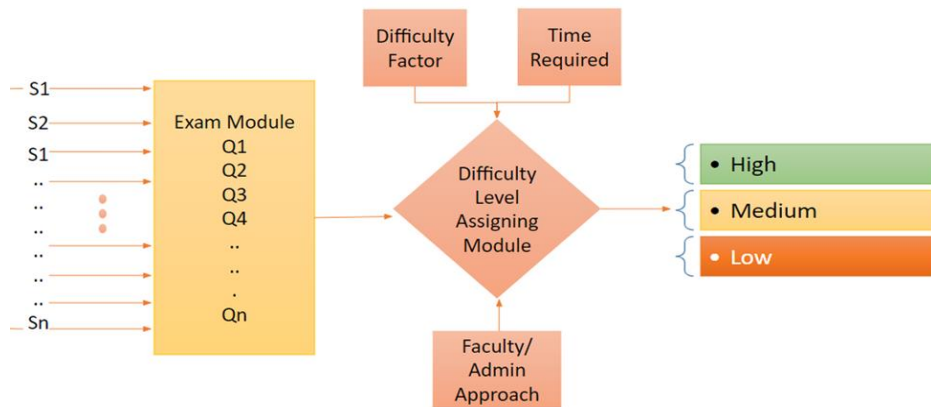


Fig 1. Question difficulty calculation module

database.

b. Question Selection Mechanism

When a candidate gives a response to a question, a feedback signal is sent to a question selector module, where the difficulty level of the next question to be assigned is decided, after determining the difficulty lever a random question belonging to that particular difficulty level is fetched from the question bank. This cycle continues for the number of questions assigned by the admin. For new questions, the difficulty level is set by the exam administrator.

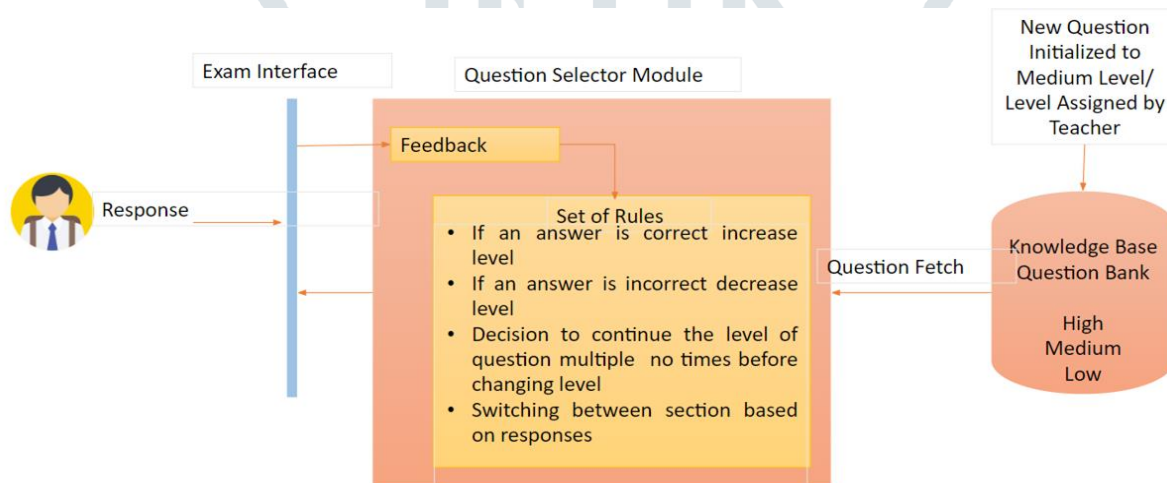


Fig 2. question selector mechanism

IV. ORDINAL LOGISTIC REGRESSION

Ordinal logistic regression (also known as ordinal regression) is an addition to binomial logistic regression. The Ordinal logistic regression is used here because it can be used to predict a dependent variable with help of some independent variables, here, the difficulty level is the dependent variable and the independent variables are the time required to solve a question and the correctness of the response to a question. So it is beneficial for assigning difficulty level to a particular question during every response by a candidate.[7][3]

• **Mathematical formulation:**

Suppose the proportions of questions of the statistical population which would be answered correctly and took less time to be p_1 , correctly and took more time to be p_2 , incorrectly and took less time to be p_3 and, incorrectly and took more time to be p_4 . Then the logarithms of the odds of classifying in certain ways are:

$$\text{correctly and took less time, } \log \frac{p_1}{p_2+p_3+p_4}, \text{ very easy}$$

$$\text{correctly and took less time or correctly and took}$$

$$\text{more time, } \log \frac{p_1 + p_2}{p_3 + p_4}, \text{ easy}$$

$$\text{correctly and took less time or correctly and took}$$

more time or incorrectly and took less time,
 $\log \frac{p1 + p2 + p3}{p4}$, moderate
 correctly and took less time or correctly and took
 more time or incorrectly and took less time or
 incorrectly and took more time ,
 $\log p1 + p2 + p3 + p4$, difficult

with these equations, the classification of the questions is done into very easy, easy, moderate, and difficult questions.

V. DATA COLLECTION

Some data sets were required for initial experimental purposes, it was obtained through an online survey platform named SurveyMonkey. A quiz was created and was circulated over various social media platforms, lots of people attempted these questions and a database was created based upon the responses. These responses were useful to set the initial difficulty levels for the computerized adaptive assessment module.

VI. RESULT

Question number	Difficulty level according to teacher	Difficulty level as per the system
1	easy	0.22 (easy)
2	easy	1.45(moderate)
3	easy	2.52(difficult)
4	moderate	0.16(easy)
5	moderate	3.21(difficult)
6	moderate	2.35(difficult)
7	difficult	1.62(moderate)
8	difficult	0.73(easy)
9	difficult	1.84(moderate)
10	difficult	2.59(difficult)

Table 1. Comparison between difficulty levels assumed by teachers and difficulty levels predicted by the system

The above table is the comparison between the estimation of the difficulty levels of the questions done by the teachers and the prediction of the system based on the responses by 45 students. The figures showed in the third column i.e, "Difficulty level as per the system" is those determined by the calculations. If the rating is >1 then it is marked easy, if the rating is ≤ 1 and >2 then it is marked as moderate and if it is ≤ 2 then it is marked as difficult.

While comparing the two columns we can see that the teacher failed to estimate the difficulty level of some questions. The questions which the teacher assumed to be easy turned out to be difficult for students also the questions which the teacher considered to be difficult, was a very easy one for students. So this system could be very beneficial from the student's point of view and could help to bridge the gap between students and teacher's thinking styles.

VII. CONCLUSION AND FUTURE WORK

Computerized adaptive assessments can be proven to be highly beneficial for school students, where the results can help the teachers to analyze a particular student and find those students who need some special attention in some particular topic such that those students can excel in those weaker areas next time. Also, this test is highly beneficial to understand the degree of attention required for a particular student. The future works for this platform will include adding more broad type of questions like oral type, long answer type, short answer type, and fill in the blanks type questions to this platform, which can be achieved by various text analysis and speech analysis techniques.

REFERENCES

- [1] Automatic Classification of Question Difficulty Level: Teachers' Estimation vs. Students' Perception. Elena Verd ú P érez, Luisa M. Regueras Santos, María Jesús Verd ú P érez, Juan Pablo de Castro Fernández, Ricardo García Martín. 2012
- [2] Course Knowledge assessment tool using Python. Parkavi A. 2016 Chen, W. et al. (Eds.) (2016). Proceedings of the 24 th International Conference on Computers in Education. India: Asia-Pacific Society for Computers in Education

- [3] A Review of Regression Procedures for Randomized Response Data, Including Univariate and Multi-variate Logistic Regression, the Proportional Odds Model and Item Response Model, and elf-Protective Responses. M.J.L.F. Cruyff, U. Bocknholt P.G.M. van der Heijden and L.E. Frank. 2016
- [4] Rating Difficulty Level of Questions(Patent). Donna K. Byron, Petersham, Suzanne L. Estrada, Boca Raton, Alexander Pikovsky. 2017
- [5] Ching, Gwo, Chien, Chih (2012).Effects on learners' performance of using selected and open network resources in a problem-based learning activity, *British Journal of Educational Technology*, Volume 43, Issue 4, Pages:606–623.
- [6] W. Zhou, “Teachers’ estimation of item difficulty: What contributes to their accuracy?,” *Proc. 31st Ann. Meet. North Am. Chapter Int. Group Psychol. Math. Educ.*, 2009, pp. 261-264.
- [7] *Regression Modeling Strategies With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis* Author Frank E. Harrell, Jr., Online ISBN 978-3-319-19425-7
- [8] *Development and Introduction of a Computer Adaptive Graduate Record Examinations General Test*. Craig N. Mills, American Institute of Certified Public Accountants. *Innovations in Computerized Assessment*, Chapter 6. ISBN 9781410602527.

