

# Research on Algorithm of MOOC Automatic Rolling System Based on Big Data Analysis

Yong Luo<sup>1,2,\*</sup>, Xuchong Liu<sup>2</sup>, Wei Zhao<sup>2</sup> and Xiao Xiao<sup>3</sup>

<sup>1</sup>College of Science, National University Of Defense Technology, Changsha, China

<sup>2</sup>Hunan Provincial Key Laboratory of Network Investigational Technology, Changsha China

<sup>3</sup>Higher Education Press, Beijing, China

\*Corresponding author

**Abstract**—At present, the credibility of the MOOC certificate is facing serious questions. Unscientific testing is one of the main reasons. This article analyzes the characteristics of MOOC test results data. Based on the distribution it obeys, an algorithm for automatically generating test papers is constructed. According to the probability of normal distribution, determine the number of questions, and a standardized test template was designed. Using the knowledge point parameters, the knowledge coverage of the test paper can be guaranteed. This algorithm can improve the reliability of MOOC testing. It can increase the social acceptance of MOOC certificates, promote the development of MOOC.

**Keywords**—MOOC; rolling system; Gaussian distribution; standard template

## I. INTRODUCTION

MOOC is the large open online courses. Courses are provided by internationally renowned university and shared through the network [1]. It enables global sharing of quality course. Coursera, Udacity, edX are currently the three largest providers in the world.

Its philosophy is to enable global learners to share quality resources through information technology and web technologies. [2] MOOC have a complete set of teaching modes. Including registration, watching video lectures, quizzes, assignments, discussions, exams, graduation, certificates and other processes. These teaching models to learners brought a new experience and give chance to get involved in high education. At the same time, through sharing the resources, the college courses can also be improved.

Although it has been rapid development and recognition, but there are still many controversies. From the history of online education, changes in learning resources promote the revolution in learning styles. How to use online course resources scientifically is a common concern. MOOC is currently facing such problems as high dropout rates, low resource utilization and lack of an effective profit model. MOOC courses often enroll tens of thousands of students, but often the last certificate students only 1%.

The MOOC test system is currently an important application model in universities. It generally uses a fixed test paper to test, and the system just orders the answers to the questions. However, there are few studies on the algorithm of automatic coiling system for MOOC evaluation. Some recent studies include the following. Formanek (2018) [1] analysed

and studied motivations of college students studying MOOC courses. Lloyd et al. (2017) [2] calculated and analysed MOOC course learning behaviours and learning performance data. Andrew (2017) [3] based on the characteristics of chemistry courses, studies how to enable more effective communication among learners and improve their learning in MOOC. Alario-Hoyos (2016) [4] analysed MOOC learning behaviours and its effectiveness as a means of assessing MOOC. Hone et al. (2016) [5] analysed the factors affecting the retention of MOOC learning through the questionnaire. Mackness et al. (2015) [6] analysed the prospects and challenges of MOOC development. Margaryan et al. (2015) [7] analysed MOOC teaching quality and its influencing factors. Phan (2016) [8] analysed students' performance in participating in the MOOC course. And do a comparative study of classroom learning and online learning. Bart Pursel et al. (2016) [9] studied MOOC learners' demographic data, learning behaviours and course interaction data. Stich (2017) [10] analysed the MOOC in the United States and found that the platform blindly pursued the number of courses but neglected the service to learners.

This article analyzes the problems faced by the existing evaluation system. Based on statistical analysis of the test, a scientific test paper generation algorithm is designed. The study includes the division of knowledge points, the design of test questions and the test model generation model. We have completely constructed the MOOC test system. This study can improve the scientific and effectiveness of the test. It can also promote the further promotion of MOOC and be accepted by more people.

## II. PROBLEMS WITH MOOC TEST SYSTEM

Because the core goal of MOOC is to learn knowledge, practice is an effective means to promote learning. MOOC platform setup exercises are necessary. In MOOC practice, the most useful resource for classroom teaching is testing. Of course, there are some problems in the current practice module, mainly reflected in two aspects. Test type is limited by the score and input mode; There is a lack of uniform standards for practice resources, and the sharing of resources is limited. Of course, many researchers are also trying to convert subjective questions into objective questions, thereby increasing the types of exercises.

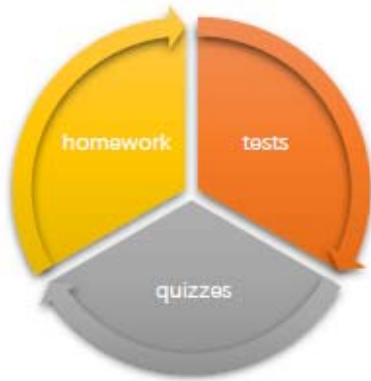


FIGURE I. THREE TYPES OF EXERCISES

At present, the MOOC platform mainly includes three types of exercises: homework, quizzes, and tests. They correspond to the three steps of learning. Homework is an exercise that learners must complete after watching a video, mainly to consolidate knowledge. The quiz test is an exercise that must be completed during video learning. They are often designed not to be skipped, and learners can only continue after completing these exercises. The course exam is a necessary condition for studying assessment and certificate of completion and is also the main mode of student evaluation. This article focuses on the test roll system.

Most of the quizzes in MOOC use automatic machine review methods. With the aid of information technology, the MOOC course quiz can also allow students to make multiple attempts. The main problems of the existing test system are as follows.

- (1) The test questions are solidified, so students can easily cheat, and the credibility of the test is not high.
- (2) There is a lack of quantitative analysis of knowledge coverage. The difficulty of the questions cannot be accurately measured.
- (3) There are no technical standards for the examination resource, and the scalability and versatility are poor.
- (4) There are too few question types and there is difficulty in subjective test.

### III. TEST DISTRIBUTION TEST

In order to design a roll algorithm, we need to check the distribution of test scores. Assume that the tests are independent and not affected by each other, so you can guess that the scores are normally distributed.

Let the overall be  $\eta$ . The score of a certain MOOC test is  $\eta_1 \eta_2 \dots \eta_m$ . Then  $\eta_1 \eta_2 \dots \eta_m$  is a sample of  $\eta$ . This sample is used to test whether  $\eta$  obeys a normal distribution. We used the Pearson  $\chi^2$  test to test whether the scores obey the normal distribution.

The above problem was transformed into the following hypothesis test.

$$H_0 : F(x) = F_0(x), H_1 : F(x) \neq F_0(x)$$

Step 1: The mean  $\bar{\eta}$  and variance  $\bar{R}$  of  $\eta$  are estimated using maximum likelihood estimation.

Step 2: Divide the interval to complete the correlation. As shown in the table below

TABLE I. DATA PREPROCESSING

Number	Interval	Frequency	Probability
1	$[a_0, a_1)$	$n_1$	$p_1$
2	$[a_1, a_2)$	$n_2$	$p_2$
...			...
$l$	$[a_{l-1}, a_l)$	$n_l$	$p_l$

Among them,  $p_i$  is calculated from  $n_i$  and  $n$ .

$$p_i = \int_{a_{i-1}}^{a_i} \frac{1}{\sqrt{2\pi R}} e^{-\frac{(\eta-\bar{\eta})^2}{2R^2}} d\eta$$

Step 3: Construct statistics  $\psi = \sum_{i=1}^l \frac{(n_i - np_i)^2}{np_i}$ . Since

there are two parameters estimated here, Pearson's theorem is used to know that the random variables  $\psi$  take the  $\chi^2$  distribution. The degree of freedom is  $l-3$ . The density function is  $\bar{p}(x)$ .

Step 4: Given confidence level  $1-\alpha(0 < \alpha < 1)$ . Counting points  $\chi^2_{(l-3)}(\alpha)$ .

$$\int_{\chi^2_{(l-3)}(\alpha)}^{+\infty} \bar{p}(x) dx = \alpha$$

Calculate  $\chi^2_{(l-3)}(\alpha)$  using the table lookup method.

Step 5: Calculate the value of  $\psi$  and compare it with  $\chi^2_{(l-3)}(\alpha)$ . If  $\psi > \chi^2_{(l-3)}(\alpha)$ , reject  $H_0$  at the level of significance  $\alpha$ , otherwise accept  $H_0$ .

### IV. TEST DESIGN SPECIFICATION

Learning resources will be integrated according to knowledge points. Learners can choose resources for knowledge. Of course, the learner can also choose all the course resources of a lecturer to learn. These resource types all require a unified template and technical standards.

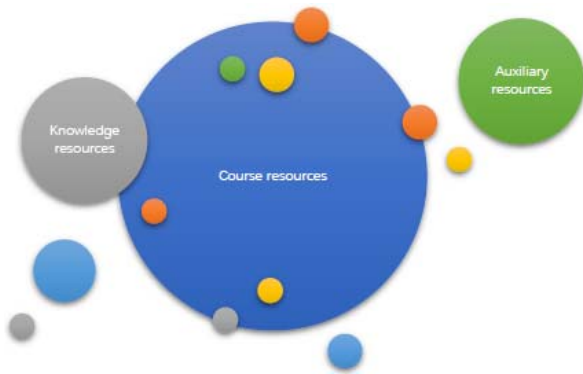


FIGURE II. COURSE RESOURCES AND THEIR CLASSIFICATION

Knowledge resources are divided into concepts and theorems. Each knowledge point resource attribute should contain knowledge point coordinates, knowledge point classification, resource type, resource provider, and so on. Concepts are definition descriptions and definitions. Theorem includes theorem description, proof and explanation. The general resource types are video, PPT, and test questions. Testing is the most used resource in universities and is also the basis for curriculum evaluation. Question design should have the following elements:

- (1) Question knowledge coordinates: Course comprehensive test questions should contain at least 2 or more knowledge point coordinates;
- (2) Test type: real-time test, quiz test, unit test, comprehensive course test;
- (3) Question type: judgment, radio, multiple choice, answer;
- (4) The difficulty level of the test questions: A (hardest), B (harder), C (medium), D (easily), and E (easiest);

TABLE II. QUESTION INFORMATION

Ticket number	90002010208001			
Knowledge point	010201	010203		
Weights	3	7		
Question type	Difficulty	Test type	Related examples	Learning major
Question Remarks				

V. AUTOMATIC ROLLING SYSTEM DESIGN

At present, there is no perfect test bank system for MOOC courses. All tests are solidified in the system at the beginning of the course construction. Moreover, there was almost no maintenance and updating in the later period, which led to a continuous decline in the role of testing. Answers for many popular course questions can be found on the web. Causes the test to completely lose its meaning. Ultimately affect the authority of the course certificate.

If we can establish a scientific and effective automatic coiling system, we can greatly improve the reliability of online testing. Automatic roll-up systems can automatically generate course tests. Use the normal distribution method to calculate the percentage of each difficulty test question in the paper.

According to the theory of education, the model should cover the coverage of knowledge points. The factors considered include the expected value of the performance, the total number of questions, the knowledge coverage rate, the type of questions and the level requirements for the assessment. The expected value of a set of papers determines the overall difficulty of the paper. Student exams are mutually independent events. Therefore, it is reasonable to assume that the MOOC assessment results obey the normal distribution. By analyzing and calculating existing test scores, we find that they are in a normal distribution.

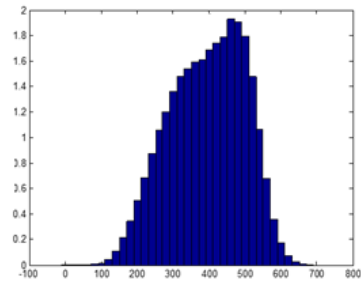


FIGURE III. MOOC TEST STANDARD SCORE DISTRIBUTION

As shown in Figure 3, we first standardized test scores. The standard formula is:

$$T = 100 \cdot \frac{X - a}{\sigma} + 500$$

To improve the resolution of the data, the data is magnified and panned. 100 here to enlarge parameters, 500 for the translation parameters. Through this formula can be seen, equivalent points can be decimal, such a score more accurate.

The expected value of a well-designed paper should be roughly equivalent to the actual average. The proportion of test scores for various difficulty in the test paper should also be roughly normal distribution. The proportion of test questions in various difficulty levels should also be a normal distribution. This can achieve the desired result by changing the proportion of various difficulty questions.

Set the test volume for each difficulty to normal distribution.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Where  $\mu$  is its expectation and  $\sigma$  is its mean square error. According to the theory of probability, for a normal random variable, the probability that its value belongs to interval  $[\mu - 5\sigma, \mu + 5\sigma]$  is 99.97%.

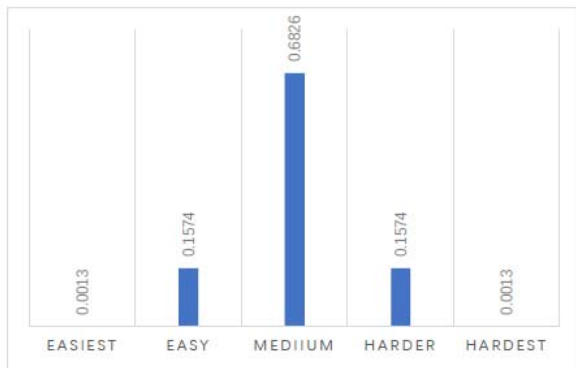


FIGURE IV. TEST QUESTIONS ACCOUNT FOR THE PROPORTION OF TESTS

Scale and shrink interval  $[\mu - 5\sigma, \mu + 5\sigma]$  to  $[0,100]$ . Divide the  $[0,100]$  interval into five equal parts. That is  $[0,20]$ ,  $[20,40]$ ,  $[40,60]$ ,  $[60,80]$ ,  $[80,100]$ . Calculate the probability that the random variable falls within the interval. Let it correspond to A (hardest), B (harder), C (medium), D (easily), and E (easiest) in the degree of difficulty of the questions. And this probability is taken as the proportion of the scores that correspond to the difficulty test questions in the test paper. According to the proportion of the questions, the system selects the corresponding questions from the question bank to generate test papers.

## VI. CONCLUSION

The current MOOC test is highly fixed. Students can easily be cheating. Even the answer to the course can be found on the web. As a result, the credibility of the MOOC certificate was questioned. This article analyzes the mathematics of test scores. Based on this rule, an automatic rolling system was designed. The test questions are automatically matched by the computer according to the set expectations. It also includes the design of test templates, which greatly improves the scalability of the questions.

## REFERENCES

- [1] Formanek, M., Wenger, M., Buxner, S., Impey, C.D. (2018). Motivational Differences between MOOC and Undergraduate Astronomy Students. *American Astronomical Society*, 231.
- [2] Lloyd, P. R. (2017). Participation patterns in a massive open online course (MOOC) about statistics. *British Journal of Educational Technology*, 48(6), 1295–1304.
- [3] Andrew, A. (2017). The nature and level of learner–learner interaction in a chemistry massive open online course (MOOC). *Journal of Computing in Higher Education*, 29(3), 411–431.
- [4] Alario-Hoyos, C., Muñoz-Merino, P. J., Pérez-Sanagustín, M., et al. (2016). Who are the top contributors in a MOOC? Relating participants' performance and contributions. *Journal of Computer Assisted learning*, 32(3), 232–243.
- [5] Hone, K. S., & El Said, G. R. (2016). Exploring the factors affecting MOOC retention: A survey study. *Computers & Education*, 98, 157–168.
- [6] Mackness, J., & Bell, F. (2015). Rhizo14: A rhizomatic learning cMOOC in sunlight and in shade. *Open Praxis*, 7(1), 25–38.
- [7] Margaryan, A., Bianco, M., & Littlejohn, A. (2015). Instructional quality of massive open online courses (MOOCs). *Computers & Education*, 80, 77–83.

- [8] Phan, T., McNeil, S. G., & Robin, B. R. (2016). Students' patterns of engagement and course performance in a massive open online course. *Computers & Education*, 95, 36–44.
- [9] Pursel, B. K., Zhang, L., Jablokow, K. W., Choi, G. W., & Velegol, D. (2016). Understanding MOOC students: motivations and behaviours indicative of MOOC completion. *Journal of Computer Assisted learning*, 32(3), 202–217.
- [10] Stich, A. E., & Reeves, T. D. (2017). Massive open online courses and underserved students in the United States. *The Internet and Higher Education*, 32, 58–71.