Integrating IRT to Clustering Student's Ability with K-Means

Wen-Chih Chang
Chung Hua University, Taiwan
earnest@chu.edu.tw

Sheng-Lin Chen
Chung Hua University, Taiwan
Changzero1985@yahoo.com.tw

Mao-Fan Li
Chung Hua University, Taiwan
M09610030@chu.edu.tw

Jui-Yu Chiu
Chung Hua University, Taiwan
Lances55@hotmail.com

Abstract

Examination plays a role to judge learner's learning behavior and achievement in evaluation. In most cases, good grade means good learner. Teachers do not realize what learners know and how much they understand. Learners with poor grades are becoming giving up them easily. Modern evaluation, diagnoses students with learning ability not grade. There are two assumptions. First, the difficulty level of materials is suitable for the students. Second, the difficulty level of question matches the teaching material. The main purpose is diagnosing the student's ability. This research calculates the student's ability from online-test system with Item Response Theory (IRT). We integrate K-means to cluster learner's ability which is calculated from item response theory. Teachers can modify the learning material adaptively and teach students in accordance with their aptitude in their courses.

Keywords: Student's ability, Item Response Theory, K-means

Introduction

Learning becomes very easy in this modem society. Everyone can enjoy the learning everywhere anytime. As long as the convenience of the internet, distance is not the problem to learning. Internet plays an important role in learning in the future. Everyone can get information which is they want. Teacher can upload materials on the internet.

In the traditional teaching, most knowledge and concepts can be compelled directly to students. The status of student is measured by the test scores before, but teachers never think about the ability of the student. Some students have the same score; they do not understand the same information in the learning material. However, this issue is not taken seriously.

Most of the test analysis only discussed with difficulty and discrimination in traditional test, but the pseudo-chance parameter is not. Students might not aware of their answer in test when answering the questions. In this state, they did not suppose to know these examinations. In this way, it is not easy to determine each student's study situation. Therefore, these results of the test analysis can not help teachers to enhance the effectiveness of teaching materials in class.

This study by the internet popularity and wireless networks are becoming more and more stable growth, teacher can edit the examination paper through the system. After finish the test, teacher can used Item Response Theory (IRT) to analyzed and understood every student's learning ability. For the analysis results, teacher can adjust the direction of teaching material and understand difficult points with the student in class. With a complete analysis of the structure, this system can help user to know the study situation of the student in order to achieve the best teaching results.

Related Work

Classical Test Theory [5] and Modern Test Theory [6] are important methods which can quantify to abstract data and analysis the student's information.

Classical test theory is mainly based on test score. Modern Test Theory is based on the item response theory. In the two theories, the difficulty and the discrimination are important attribute to examine. But the most difference between the two theories is considering the impact of questions on the parameters. However, classical test theory can not think about this issue [1].

Item Response Theory (IRT) is one of the ways in Modern Test Theory. To use the chart showing the test characteristic curve from difficulty and discrimination. Min-Ning Yu [9] mentioned that the test characteristic curve can display the student's ability and probability of correct response from learning [8]. These characteristic curves will be change by different parameters. The math function, which predicts of student's ability, and enhances the effectiveness of learning, is build by the characteristic

978-0-7695-3873-0/09 $29.00 © 2009 IEEE

curves. There are three functions: one-parameter logistic model (the difficulty), two-parameter logistic model (the difficulty, the item discrimination), three-parameter logistic model (the difficulty, the item discrimination and the pseudo-chance parameter). Though Item Response Item, students can understand their own ability, and the student's ability can not be affected by time and amount. Furthermore, item response theory can collocate with other theories to produce different effect in the result. For example: Item response theory applies to analysis the test attribute and tester behavior with the student-problem chart(S-P) [2].

Clustering is classified into different clusters with scattered information. User can observe differences in the host conveniently from every distribution of the clusters. In addition, it is used at data miming which can be predicted the behavior of information [4][7]. At a large number of clustering methods, K-means is often used method. By several centers of mobile, user can look for meet the minimum distance between data points. Finally, achieve a balance between the various clusters. Because the cost is very low, it often combines with different theories [3] [10].

Main Method

There are two ways which can help us to understand the student's ability. One way is to observe the performance of students in class, the other way is test score. Item response theory is used to forecast the student's ability from answering questions. By the item response theory, teachers can analysis and understand that how much student learned, and provide helpful remedial information to students and instruction hint to teachers in class.

Research Course and Item Analysis

The subjects of this research are all of the sophomore students in department of information management in one of the universities in Taiwan. The research course is "Data Communications for Business". The exam has sixty multiple choice items. In order to use the item response theory formula for forecast the student's ability after the test, we analyses the difficulty and the item discrimination in the test first.

Difficulty (b_i): It is item difficulty value. "i" is the item number in test. If the b_i value is higher, this item is easier. Formula 1 shows the difficulty function, P_H is the correct rate of the high score group, and P_L is the correct rate of the low score group. The high score group and low score group are get 27% for all of the students.

$$b_i = \frac{(P_H + P_L)}{2} \qquad \text{Formula 1}$$

Item Discrimination (a_i): It is item discrimination value. If a_i value is higher, this item is easier to identify if students realize it. Formula 2 is the function of item discrimination.

$$a_i = P_H - P_L \qquad \text{Formula 2}$$

Forecast the Student's Ability

This section is the main method in the research. There are three functions discussed the difficulty (b_i), the item discrimination (a_i) and the pseudo-chance parameter (c_i) in Item Response Theory.

(1) One-Parameter Logistic Model: It is forecast the student's ability from difficulty (b_i). It is think about difficulty value could be affect the student's ability. In Formula 3, $P_i(\theta)$ is the correct rate of the i item which answered by the student of the ability value θ. This value range between 0 to 1. The value of e is 2.718. This model does not apply a few items which are too easy or too difficult.

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \qquad \text{Formula 3}$$

(2) Two-Parameter Logistic Model: In Formula 4, a_i is the discrimination of the i[th] item. The learning ability could be calculated by the difficulty and the item discrimination in this model. If two students have the same number of wrong question with different items from this test, the $P_i(\theta)$ would be not the same. The ability of the student can be more objective.

$$P_i(\theta) = \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}} \qquad \text{Formula 4}$$

(3) Three-Parameter Logistic Model: In Formula 5, pseudo-chance parameter (c_i) is put into consideration, because every student might guess the item in the test.

$$P_i(\theta) = (1 - c_i) + c_i * \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}} \qquad \text{Formula 5}$$

This research used the intercept to count the pseudo-chance parameter (c_i) from the characteristic curve. In order to get more accurate intercept, the test information is used in this study. In characteristic curve, there is better reliability near the vertex. Formula 6 is the function of test information characteristic curve. $I_i(\theta)$ is the test information at the i[th] item. $P_i(\theta)$ is the correct rate of the i[th] item which from the ability value θ of the student. $Q_i(\theta) = 1 - P_i(\theta)$ is the incorrect rate. a_i is the discrimination value.

$$I_i(\theta) = a_i{}^2 P_i(\theta) Q_i(\theta) \qquad \text{Formula 6}$$

In this research, the effective and reliability range is 0.5 units near the vertex. Figure 1 is characteristic curve of the 60th item. For example: when the ability value is 0.8 which can find the reliability test information. So we get 0.3 to 1.3 ranges to count the pseudo-chance parameter which value (c_{60}) is 0.0227. Formula 7 is the function of student's ability.

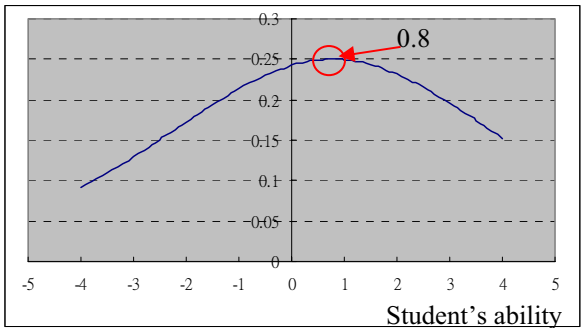

Figure 1. Test information characteristic curve of the 60th item

Table 1. Count to student's ability

item	u	P	Q	a(u-P)	a*a(PQ)	Δθs	θs+1
1	1	0.5	0.5	0	0	2.24522	3.24522
2	1	0.50103	0.49897	0.04536	0.00207		
3	1	0.50103	0.49897	0.04536	0.00207		
:	:	:	:	:	:		
:	:	:	:	:	:		
:	:	:	:	:	:		
58	1	0.56472	0.43528	0.277	0.09954		
59	1	0.56573	0.43427	0.31583	0.12995		
60	1	0.51549	0.48451	0.04405	0.00206		
				5.79764	2.58221		

$$\theta_{s+1} = \theta_s + \frac{\sum_{i=1}^{N} a_i [u_i - P_i(\theta_s)]}{\sum_{i=1}^{N} a_i^2 P_i(\theta_s) Q_i(\theta_s)} \qquad \text{Formula 7}$$

Table 1 is the real example in the research and 60 items in the test. u is the student's performance in the title, 1 means correct and 0 means incorrect. $\Delta\theta_s$ is variation in the volume of student's ability. If this value is lower, the student's ability will be more accurate. $\Delta\theta_s$ must lower than 0.001 in our research. θ_{s+1} predicts the value of student's ability.

Results

In the research, we used two experiments to test the validity of three parameter Logistic Models in Item Response Theory. First experiment is clustering effect of the three models tested in the result. Second experiment is grouping analysis of the effectiveness in class.

Clustering Effect of the Three Models Tested in the result

Table 2, Table 3 and Table 4 are the experimental data after K-means clustering. In Table 2, we can know that centers of each cluster group are too close to each other in One-Parameter Logistic Model. But this result does not haves obvious effect. In Table 3 and Table 4, centers of each cluster group are not intensive and easy to see the difference in clustering group. These results are in line with the research needs.

Table 2. The result of One-Parameter Logistic Model

Group	1	2	3	4	5	6	7	8	9	10
Center	4.16	4.87	3.74	2.80	3.29	1.18	2.08	1.64	2.50	0.85
Numbers	1	4	2	6	7	6	7	7	4	2

Table 3. The result of Two-Parameter Logistic Model

Group	1	2	3	4	5	6	7	8	9	10
Center	1.03	26.13	15.86	5.11	6.09	7.73	9.91	2.44	3.79	12.47
Numbers	7	2	1	7	2	5	3	10	8	1

Table 4. The result of Three-Parameter Logistic Model

Group	1	2	3	4	5	6	7	8	9	10
Center	0.50	25.85	10.06	4.80	5.79	7.65	15.57	2.05	3.45	12.18
Numbers	7	2	2	7	2	6	1	10	8	1

Grouping Analysis of the Effectiveness in class

In this section, there are one hundred and sixteen students participate this experiment. Those students were divided into two groups and implement pre-test and post-test. One is the experimental group and the other is the control group. After the pre-test, we use the IRT function to predict the student's ability of experimental group and divide them into eleven groups, who have a forty minutes group discussion. The control group did not have group and discussion.

After the post-test, we compared the two groups with student's ability. In experiment group, the average student's ability of pre-test is 3.84, and the average student's ability of post-test is 5.97. In control group, the average student's ability of pre-test is 2.16, and the average student's ability of post-test is 2.4. The reliability of pre-test and post-test is 0.86 and 0.76 with KR-21 (Kuder-Richardson reliability). In both experimental tests, the test result of the experiment group is more progressive than control group. We found that the group members did not enjoy their discussion each groups. This state must be observed and improved to us.

Conclusion

In this research, student's ability is been predicted by Item Response Theory, and use K-means to clustering the student's ability. It can help teachers to adjust teaching curriculum and materials. The test is not a tool to measure the student, but help teachers understand the student's ability. The classification information can be applied in remedial course to enhance the teaching effectiveness.

We use one-dimension and K-means in our clustering method, we hope to increase more dimensions in the future, for example: increase the rate of student's attendance or the performance in class. At the same time, we may choose different clustering method compared with the K-means, for example: Fuzzy C-means. Hope that by grouping different method and effective method of data analysis to get more accurate result.

Acknowledgements. We would like to thank National Council. This research was supported in part by a grant from NCS-98-2511-s-216-002, Taiwan, Republic of China.

Reference

[1] Chih-Ming Chen & Yi-Yun Chen, 2005, Learning Performance Assessment Approach Using Learning Portfolio for E-learning Systems, Proceedings of the Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05), pp557-561.

[2] Deng-Jyi Chen & Yu-Wei Chan , 2006, The analysis of the test item attributes and the testee' response behaviors based on the Item Response Theory (IRT) and Student-Problem (S-P) Chart ― using multimedia English test questions as example, A Thesis Submitted to Degree Program of E-Learning College of Science National Chiao Tung University in partial Fulfillment of the Requirements for the Degree of Master in Degree Program of E-Learning.

[3] Du Jin, Zheng Qinghua, Ding Jiao, Gong Zhiyong, 2006, A Method for Learner Grouping Based on Personality Clustering, Proceedings of the 10th International Conference on Computer Supported Cooperative Work(CSCWD 2006) pp.1420-1425.

[4] Everitt, B. S.(1993). Cluster Analysis. New York: John Wiley & Sons Inc.

[5] Gullikson, H. (1987). Theory of mental tests. Hillsdale, NJ: Lawrence Erlbaum Associates.

[6] Hambleton, R. K., & Swaminathan, H. (1985). Item response theory: Principles and applications. Boston, MA: Kluwer-Nijhoff.

[7] I-Cheng Yeh，2006，Data Mining ： Processes. Model, and Practices (Third Edition)，Unpublished.

[8] Min-Ning Yu, [On-line]. Available: http://www.edutest.com.tw/e-irt/index.htm

[9] Min-Ning Yu, 1997, Testing and assessment in education, Taiwan: psychology.

[10] Shian-Shyong Tseng, Pauray S.M Tsai, Tung-Shing Su, Chiu-Jung Tseng& Ching-Yao Wang, 2006, Data Mining, Flag Publishing in Taiwan.