Design and Evaluation of Training System for Numerical Calculation Using Questions in SPI2

Shin'ichi Tsumori[1] and Kazunori Nishino[2]

[1] Kyushu Junior College of Kinki University, Japan
tsumori@kjc.kindai.ac.jp
[2] Kyushu Institute of Technology, Japan
nishino@lai.kyutech.ac.jp

Abstract. We are developing the training system for numerical calculation aiming at improving calculation ability. There are two main purposes of realizing this system. One is to increase students' motivation to study mathematics by using the questions in SPI2 adopted by many companies as employment examinations. The other is to support a student's learning efficiently by giving the questions according to the student's ability. In order to give an adaptive question, our system has functions to estimate each student's ability and item difficulty in the test item database. This paper reports the basic concept, the features and the experiment conducted to verify the usefulness of the system and its result.

Keywords: Training System, Numerical Calculation, SPI2, Web-Based Learning.

1 Introduction

In recent years, in many universities or colleges, the students who lack calculation ability are increasing in number, and it interferes with some lectures premised on calculation ability. In addition, since many companies give calculation tests as a part of their employment examinations nowadays, it is difficult for students with low abilities to pass them. For these reasons, many universities/colleges give remedial mathematics course to the freshmen students. However, since the difference of the calculation ability between students is large, it is difficult to raise learning effectiveness with the same curriculum. In such a case, the individualized learning which has been adapted for each student's understanding is more effective than group lessons. However, since most of arts students do not like to study mathematics generally, it is difficult for them to maintain the motivation to study using the existing mathematics materials. Instead, their motivation to study for an employment examination is comparatively high. Therefore, it is expected that the teaching materials made for employment examinations will raise their motivation for learning.

We aim to realize the web-based training system for numerical calculation using the questions of SPI2 (Synthetic Personality Inventory 2) which is widely adopted as an employment examination in Japan. This system is supposed to be used especially

M. Kurosu (Ed.): Human-Computer Interaction, Part II, HCII 2013, LNCS 8005, pp. 511–520, 2013.

in the entrance into universities/colleges or the period when a student takes an employment examination in order to acquire calculation ability. Repeated calculating is the most important until a question comes to be solved in order to gain ability. However, since there are around twenty study units in SPI2, in order to raise a learning effect in a limited time, it is important to extract the study units adequately which the student should study because she/he does not understand it well.

The system characterized by giving the questions according to the student's ability has been considered for many years. Papers [1] and [2] are featured by choosing the question which is adapted for the student's ability from the question database. They are effective methods when the test item difficulties are set to all questions beforehand, but the problem is that a load is required of a teacher. Papers [3] to [7] show the methods of generating questions automatically. Although the advantage is that they can give many questions to the students, the problem is how the validity of the difficulty of the test item in question is guaranteed.

We are developing the training system for numerical calculation equipped with two functions, ability analysis and a calculation training[8]. One of the features of the system is to avoid generating questions that will be too easy/difficult, and showing the question according to academic ability by using Item Response Theory (IRT). However, since the questions in SPI2 are added and updated frequently, there is a problem regarding how the difficulty of the test item stored in the question database is maintained. We are examining how to estimate the test item difficulty using the number of the formulas contained in the process of a numerical calculation. By using this method together with the PROX method, we think that a system which can respond to the frequent change of the contents of the question database can be realized.

This paper reports the basic concept, the features and the experiment to verify the usefulness of the system and its result.

2 System Overview

2.1 What Is SPI2?

Before we start explanation of the system developed now, SPI2 is explained briefly.

SPI is one of the aptitude tests broadly used as the employment examination of the company. It has been developed by present Recruit Management Solutions Co.,Ltd. and adopted 9,050 companies in the 2011 fiscal year. In addition, "2" of "SPI2" expresses the version number[1]. SPI2 consists of an ability test and a personality test, and the numerical calculation explained in this paper is a part of ability test. Although the range of questions of SPI2 is not exhibited, it includes around 20 study units such as probability calculation, speed calculation, and concentration calculation. SPI2 test requires the calculation ability from an elementary school to a high school. It also requires a student to answer the 30 calculation questions within 40 minutes.

In addition, although an actual ability test in SPI2 consists of multiple-choice questions, the training system which we are developing is a system which a numerical value is inputted as an answer.

---

[1] The version number of SPI as of February, 2013 is "3".

2.2 System Configuration and Operation

The system configuration is shown in Fig. 1. This system is a web-based training system and has two functions, ability analysis and calculation training. The student chooses one of the functions to use after login. The outline of two functions is explained below.

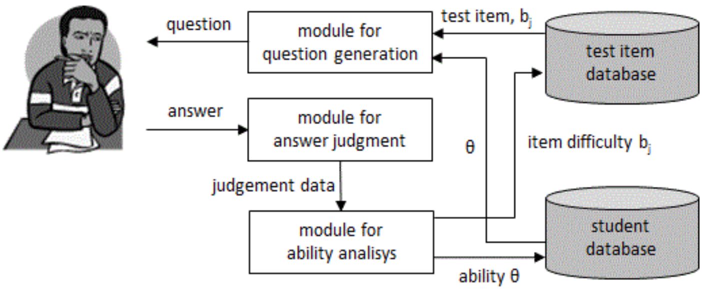

Fig. 1. Configuration of Training System for Numerical Calculation

1. Ability analysis

Ability analysis is the function to take the analysis test which consists of test items stored in the test item database. Since each of the abilities needs to become clear in advance in order to perform an adaptive training, all the students who want to use the function of calculation training must take this test. In addition, before performing ability analysis, all item difficulties in the test item database must be known. Chapter 3 explains the setting method of item difficulty.

The module for question generation gains some test items from the database, creates an ability analysis test, and gives a student the question. After a student's answer, the module for answer judgment transmits right-or-wrong information to the module for ability analysis, and the module for ability analysis estimates the student's ability θ.

We use Maximum Likelihood Estimation as the method to estimate the ability θ. It is the method of estimating θ from a viewpoint of the probability of a series of results in a test. By using the item difficulty b_j and the ability θ, the probability $P_j(\theta)$ for correct answer and the probability for incorrect answer can be calculated. Then, the probabilities that a series of student answers will take place are calculated, changing theta to various values. The largest value among the calculated values is an estimate of θ, and it is registered into the student database.

2. Calculation training

The calculation training is a function which extracts the test item suitable for the student's ability from the test item database and shows a student it. The test item has attributes and values as shown in Table.1. The contents of each attribute of a test item are as follows.

Table 1. Sample of Attributes and Values of Test Item

attribute	value
Item ID	7
Unit ID	6
Question Sentence	What percentage of salt water solution will be made if (p_1)g of water is added to (p_2)g of (p_3)% salt water solution?
Parameter set	(50, 100, 3, 2), (60, 100, 4, 2.5), (50, 200, 5, 4), …
Parent ID	11
Child ID	3, 4, 6
Item Difficulty	0.1

- Item ID
  ID of a test item
- Unit ID
  ID of a study unit
- Question sentence
  Question sentence given to a student. P_x (x=1,2,3,…) is a parameter which determines a value at the time of generating question.
- Parameter set
  Set of the numerical value given to a question sentence and the numerical value of a correct answer. There are some parameter sets in a test item, and one set is chosen at the time of generating the question. For example, when (50, 100, 3, 2) are chosen, p_1=50, p_2=100, and p_3=3 are set to a question sentence. The correct answer is 2.
- Parent ID
  IDs of test items containing the calculation process of this item
- Child ID
  IDs of test items used as the partial calculation process of this question
- Item Difficulty
  Value of test item difficulty

The module for question generation selects a test item and generates a question based on the teaching strategy. This paper does not discuss the details of a teaching strategy, but the following examples are shown as a teaching strategy.

- When a student answered a question correctly, the system gives a more difficult question using Parent ID.
- When a student answered a question incorrectly, the system gives an easier question one using Child ID.

3 Estimation of Item Difficulty

In this paper, we try to propose the method of giving a suitable question based on a student's probability of the questions stored in the test item database. We adopt the Rasch model (One-parameter logistic model) widely known for the field of IRT as a method of estimating the probability. According to Rasch model, when the item difficulty b_j and the student's ability θ is known, the probability $P_j(\theta)$ is calculated using following formula.

$$P_j(\theta) = \frac{1}{1 + e^{-(\theta - b_j)}}$$

Since $P_j(\theta)$ is a value for probability, following two points are guessed from a viewpoint of the difficulty of question.

- When $P_j(\theta)$ is close to 0, the problem is too difficult for a student.
- When $P_j(\theta)$ is close to 1, the problem is too easy for a student.

That is, the more the value of $P_j(\theta)$ approaches 0.5, the more the question will be suitable for a student. By using this view, an adaptive training system will be realized.

In order to perform ability analysis mentioned in 2.2, all item difficulties in the test item database must become clear in advance. We propose to use two methods together, the PROX method and the method of using the number of formulas contained in the calculation process in order to estimate the item difficulty b_j. These two methods are explained below.

3.1 Estimation Using PROX Method

The PROX method is one of the methods used in order to estimate student's ability θ and the item difficulty b_j required to apply Rasch model. Its feature is that the computational procedure for estimation is easier than other methods.

The example of a right-or-wrong situation where seven students answer the test, which consists of the six items, is shown in Table 2.

Table 2. Parameter estimation using the PROX method

Student ID	Question ID						θ
	1	2	3	4	5	6	
1	1	0	0	0	1	0	-0.940
2	1	1	1	0	1	0	0.940
3	0	0	1	0	0	0	-2.174
4	1	0	0	0	1	0	-0.940
5	1	0	0	0	0	1	-0.940
6	1	1	1	0	0	1	0.940
7	1	1	1	1	0	1	2.174
b_j	-2.525	0.256	-0.51	2.271	0.256	0.256	

Since a right-or-wrong situation is expressed with two values of 1 and 0, Table 2 means that the student 1 answered the question 1 and the question 5 correctly, and gave the wrong answer to other questions. The PROX method estimates θ and b_j using the logit scores calculated from Table 2 (as shown in shading). Explanation of the calculation process of the PROX method is omitted in this paper.

When θ and b_j are calculated like Table 2, the probability $P_j(\theta)$ that the student of θ= 0.5 answers the question 2 (b_j =0.256) correctly is calculated as follows.

$$P_j(\theta) = \frac{1}{1 + e^{-(0.5-0.256)}} = 0.56$$

If the difficulties of all the SPI2 items stored in the test item database are calculated in advance, when a student's ability θ becomes clear, the probability for all items will be calculated. Therefore, the system can extract the test item which is neither too easy nor too difficult for the student.

However, it is very difficult to actually set up all the item difficulties in the test item database in advance for the following two reasons.

(1) For the high updating frequency of test items in the database

The range of questions of the numerical calculation of SPI2 is wide, and also it is expanded every year. Moreover, since it is possible that various questions whose difficulties are low to high can be created, the frequency of the addition and update of the test item database becomes high. At actual schools, it is difficult to carry out frequent test for items whose difficulties are unknown.

(2) For a restriction of "assumption of local independence"

In order to apply IRT, assumption of local independence must be satisfied regarding each item of a test. Assumption of local independence means that the probability that an answer of certain item is correct does not affect the probability that answers of other items are correct. Fig. 2 shows the example of three items in the "concentration calculation" study unit. In this example, since the calculation process of item A contains the calculation process of item B, the student who cannot answer item B correctly cannot answer item A correctly. Therefore, item A and item B cannot be used for the test for ability analysis simultaneously because item A and item B are not in the relation of local independence (the relationship between item B and item C is the same). However, since it is necessary to create broadly from an easy question to a difficult one in order to give the problem according to student's ability, it is rather ordinary that items shown in Fig. 2 are simultaneously stored in the test item database. In this case, it is necessary to carry out two or more tests in order to set up the item difficulties because of assumption of local independence

3.2 Estimation Using the Number of Formulas in the Calculation Process

This section discusses how to estimate the item difficulty using the number of the formulas contained in the process of a numerical calculation.

Fig.2 shows the example of three items belonging to the "concentration calculation" and their calculation processes. In addition, all the formulas in this paper are expressed as dyadic operation.

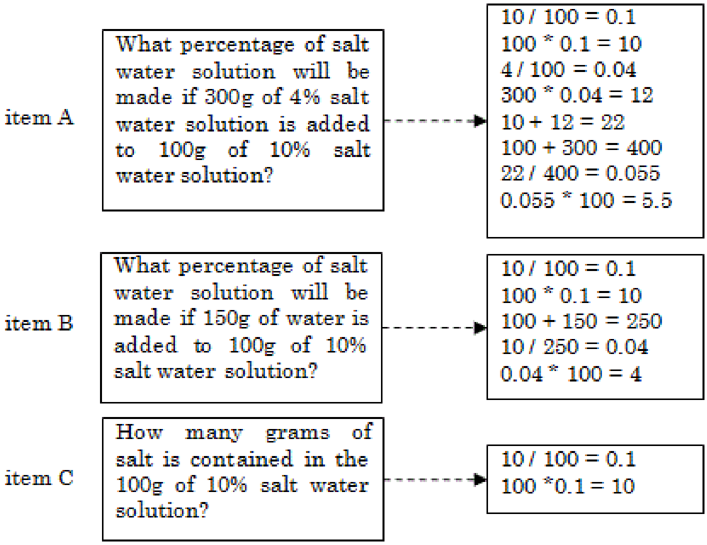

Fig. 2. Example of relationship between 3 items

As for the item C in Fig.2, the correct answer is calculated in the following process.

1. The percentage expression (concentration) is changed into the decimal expression.
2. The weight of salt is calculated by multiplying the weight and the concentration of the salt water.

As for the item B, its calculation process include one of the item C. Furthermore, some formulas are added to the item B. The calculation process of the item A include one of the item B as well. That is, the comparison of these three items guesses that the item difficulty becomes large at the order of item A, item B, and item C. For example, if the item difficulty of item B is known, the item difficulty of item A (or item C) will be estimated by getting the difference between the item difficulty of item A and item B (or item B and item C).

The number of formulas contained in the calculation process of each item of Fig. 2 are as follows.

item A: 8, item B: 5, item C: 2

Therefore, the difference between the number of formulas of item A and item B is 3. Then, we tried to make the relationship between the differences of the number of formulas and item difficulty experimentally clear. That is, we estimated the change of item difficulty corresponding to one formula in the calculation process by dividing the difference between the item difficulty of item A and item B by the difference of the numbers of their formulas of calculation processes.

The following chapter explains the experiment to perform the above-mentioned contents.

4 Evaluation of the Estimation Method for the Item Difficulty

4.1 Outline of the Experiment for Evaluation

In this chapter, we examine the relationship between the number of the formulas in the calculation process and the item difficulty. Estimation of item difficulty is performed by analyzing the answers of the test submitted by the college students using the PROX method. Under the present circumstances, the contents of the item which constitutes a test are important. For example, item A, item B and item C in Fig.2 cannot be used in the same test simultaneously because of assumption of local independence. Then, it is assumed that the difference of the item difficulty depends on only the difference of the number of formulas in the calculation process regardless of the study unit or the contents of the question.

In this paper, the relationship between the number of the formulas in the calculation process and the item difficulty is estimated in the following procedures.

1. Constitute a test using several questions from which the number of formulas in the calculation differs. Select the all questions so that the assumption of local independence may be satisfied.
2. Estimate the all item difficulties b_j of all the items in the test using the right-or-wrong situation of a student's answer.
3. Get the expression of relation between the number of formulas in the calculation process and b_j.

4.2 Method of Experiment

The experiment was conducted on the following conditions.

- Contents of the test
  - 15 questions from 4 study units ("price computation", "dealing profit and loss", "speed calculation", and "concentration calculation") are used.
  - All the questions are the numerical calculations whose correct answers are numerical values.
- Subject
  80 first graders of Kyushu Junior College of Kinki University
- Answer method
  The student accessed the Web server which stored the test using the personal computers in the classroom and answered simultaneously.
- Answer time
  Although not specified in particular, all the students finished answering the questions in about 30 minutes.

4.3 Result and Consideration

The graph which plotted the number of formulas in the calculation process of an item and the item difficulty is shown in Fig. 3. In addition, for the questions which were solved by same number of formulas, the average value of those item difficulties are the representative value. The result is that there is a strong correlation between the number of formulas in the calculation process and the item difficulty because the Pearson product-moment correlation coefficient is 0.78. The regression coefficient was 0.47, and it means that one of the formulas corresponded with 0.47 of item difficulty. Therefore, if the item difficulty of item B will be estimated at 1.00 (for example) by the PROX method, item A and item C may be estimated as follows.

- Item difficulty of item A

Since item A has 3 more formulas in its calculation process more than item B,

$$1.00 + 0.47 \times 3 = 2.41$$

- Item difficulty of item C

Since item A has 3 fewer formulas in its calculation process more than item B,

$$1.00 - 0.47 \times 3 = -0.41$$

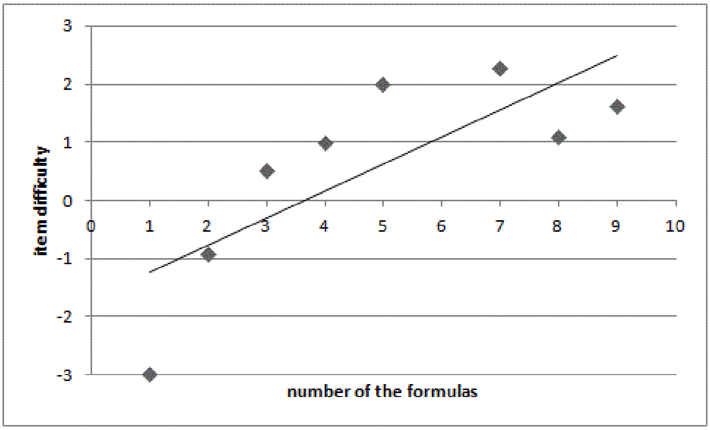

Fig. 3. Relationship between number of the formulas and item difficulty

As explained so far, it is possible that all the item difficulties can be estimated by two methods, PROX method and the method using the number of formulas contained in the calculation process. Therefore, the student's probability of each test item can be estimated with applying IRT.

5 Conclusion

We are now developing a system with the feature described so far. Moreover, we proposed the method of estimating the item difficulty using the number of formulas in the calculation process. The result of the experiment suggested the probability that the number of formulas in the calculation process may be used for estimating the item difficulty.

However, since the number of subject and the number of questions on the experiment test was not enough, we could not estimate parameters with precision. Moreover, another problem to be solved is that the rate of correct answers in an experiment test was quite lower than anticipated. Many more tests by many more students must be carried out in order to estimate item difficulty with more precision.

Acknowledgement. This work was supported by JSPS KAKENHI Grant Number 23501191. We thank to the students of Kyushu Junior College of Kinki University who cooperated in the experiment.

References

1. Huang, S.X.: A Content-Balanced Adaptive Testing Algorithm for Computer-Based Training Systems. In: Proceedings of the Third International Conference Intelligent Tutoring Systems, pp. 306–314 (1996)
2. Suganuma, A., Mine, T., Shoudai, T.: Automatic Generating Appropriate Exercises Based on Dynamic Evaluating both Students' and Questions' Levels. In: Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, pp. 1898–1903 (2002)
3. Mitkov, R., Ha, L.A.: Computer-Aided Generation of Multiple-Choice Tests. In: Proceedings of the HLT-NAACL 2003 Workshop on Building Educational Applications Using Natural Language Processing, vol. 2, pp. 17–22 (2003)
4. Holohan, E., Melia, M., McMullen, D., Pahl, C.: The Generation of E-Learning Exercise Problems from Subject Ontologies. In: Proceedings of The Sixth IEEE International Conference on Computer and Information Technology, pp. 967–969 (2006)
5. Holohan, E., Melia, M., McMullen, D., Pahl, C.: Adaptive E-Learning Content Generation based on Semantic Web Technology. In: AI-ED 2005 Workshop 3, SW-EL 2005: Applications of Semantic Web Technologies for E-Learning, pp. 29–36 (2005)
6. Gonzalez, J.A., Munoz, P.: e-status: An Automatic Web-Based Problem Generator - Applications to Statistics. Computer Applications in Engineering Education 14(2), 151–159 (2006)
7. Lazcorreta, E., Botella, F., Fernandez-Caballero, A.: Auto-Adaptive Questions in E-Learning System. In: Proceedings of the Sixth International Conference on Advanced Learning Technologies, pp. 270–274 (2006)
8. Tsumori, S., Nishino, K.: Proposal of a Numerical Calculation Exercise System for SPI2 Test Based on Academic Ability Diagnosis. Intelligent Interactive Multimedia: Systems and Services 14, 489–498 (2012)
9. Linacre, J.M.: Rasch Model Estimation. Journal of Applied Measurement 5(1), 95–110 (2004)