

Design of Computer Adaptive Test System Based on IRT

Muqing Zhan¹ and Ronghua Lu²

Jingdezhen Ceramic Institute, Jingdezhen, 333403, China

¹mushi3609@sina.com, ²lrhlucy2003@163.com

Keywords: Item Response Theory, Computer Adaptive Test, Random Extraction, Strategy of Item Selection

Abstract. Stating item response theory and computer adaptive test system, analyzing the problems existing in adaptive test system, and proposing the system design idea, taking the establishment of adaptive test system for *C language programming design* for example, it analyzes the system structure of computer adaptive test system based on IRT as well as the detailed design and realization of main functional module in details.

The adaptive test mode is always conducted surrounding the examinees' competence, and all test items are selected automatically from the item bank by the system depending on the examinees' competence: there is a difficult degree distinguished for test items, adjust the difficult degree of test items at any time according to the test accuracy, with strong pertinence, highlight the principal status of examinees and individualized demands, and enhance the validity and reliability of test, so that the test efficiency is improved. The computer adaptive test (CAT) with item response theory (IRT) as a basic theory has caused the extensive concern from each country around the world, which has been promoted gradually in various social aspects: with USA post-graduate entrance examination GRE and GMAT for the educational world; with Nurse National Committee License Test (NNCLT) for the aspect of vocational qualification authentication; with authentication test organized by Novell Company, CCNA authentication test organized by Cisco Company and MCSE adaptive test organized by Microsoft Company in the corporate world. At domestic, the research and application of IRT has caused the extensive concern from the Education Department.

Item response theory

CAT is built on the basis of modern test theory - item response theory, from the construction of item bank to the extraction of test items, and then to test paper generating and testing until to the final evaluation, it is conducted under the guidance of IRT. Therefore, CAT is identified as the greatest contribution on the test among modern test theory. For item response theory, Hambleton and Swaminathan made the following statement on it: in the test scenario, through the definition of examinee characteristic – namely peculiarity or competence, estimate the examinee's score on such characteristic (referred to as competence score), and predict or explain the item and answering condition by application of such score, explain and predict the examinee's answer. Peculiarity (competence) and item (test item) are two core concepts for IRT, and the main relationship between both items is the major connotation of IRT. The main research content and theoretical support of IRT and computer adaptive test under its guidance include: parameter estimation, test equating, item selection strategy and termination principles.

Computer adaptive test

CAT is a new kind of test with the establishment of item bank based on item response theory, automatic selection of test items by computer according to the examinee's competence as well as final estimation on the examinee's competence. It is different from traditional paper test, the presentation of CAT test items and the examinee's corresponding answer are all completed through the computer; Meanwhile, CAT is also different with general computer-based test, which is reflected in items presentation, answer input, automatic grading, performance achievement in the process of test. In addition, select the next most suitable test items automatically according to the

examinee's different answer, so as to realize the most suitable estimation on examinee's competence finally, therefore, for CAT, "The test is conducted differently from person to person".

For the application of adaptive test conducted in the earliest, it is an intelligence test research made by Binet in 1912. In the middle 60s, Lode conducted more complete overall research when working in Educational Testing Service, he thought that, the test with fixed length can't meet the demands on competence estimation for examinee with different levels, however, CAT can solve this problem: if all items selected for test can provide the most reference information aiming at each examinee's competence, then the shortening of test length (namely, reducing the quantity of items for test) shall not reduce the accurate measurement on each examinee's competence. From the theory, the test items received by each examinee shall be different test groups.

Problems existing in CAT system

Determination of parameters for item bank

This item bank is derived from problem sets of *C language programming design* used in our school, there is no ready-made data for use, it needs to set the parameters of each item, whereas the accuracy test of items is insufficient currently, the setting of parameters need to be optimized, so that the item bank can be the most reasonable.

Problems on performance of IRT system

As the extraction of test items is realized by use of IRT for CAT system, for the selection of the next most suitable test items according to the examinee's accuracy, it is involved in complex algorithm, and the complexity of time is relatively large. It needs to conduct repeated iteration when conducting IRT test, calculating the parameters of each item done again, so as to estimate the user's competence level. Therefore, it will consume a large amount of bandwidth and resource on server.

Design and realization of CAT system based on IRT

Division of system module

According to the requirements of system function, CAT test system can be divided into three modules, namely student online examination subsystem, teacher management subsystem and administrator subsystem. The system can be designed through the division of several function modules, and the relationship among various functions module is as shown in Figure 1:

Design and realization of system test module

The following function is realized by student examination subsystem:

- ① Login – this subsystem will receive the personal information input by the examinee and verify it, the legitimate examinee is allowed to enter the test system and participate in the test.
- ② Answering in limited time- the legitimate examinee receives the test, the system starts to time, and the examinee answers the items.
- ③ Submission – including the paper submitted by the examinee on his/her own within the specified time as well as the paper submitted due to the force of system.
- ④ Automatic grading – After the submission made by the examinee, the system will grade the paper automatically, and the test scores are displayed to the students.
- ⑤ Result record – insert the result obtained by the examinee into a grade table in database.

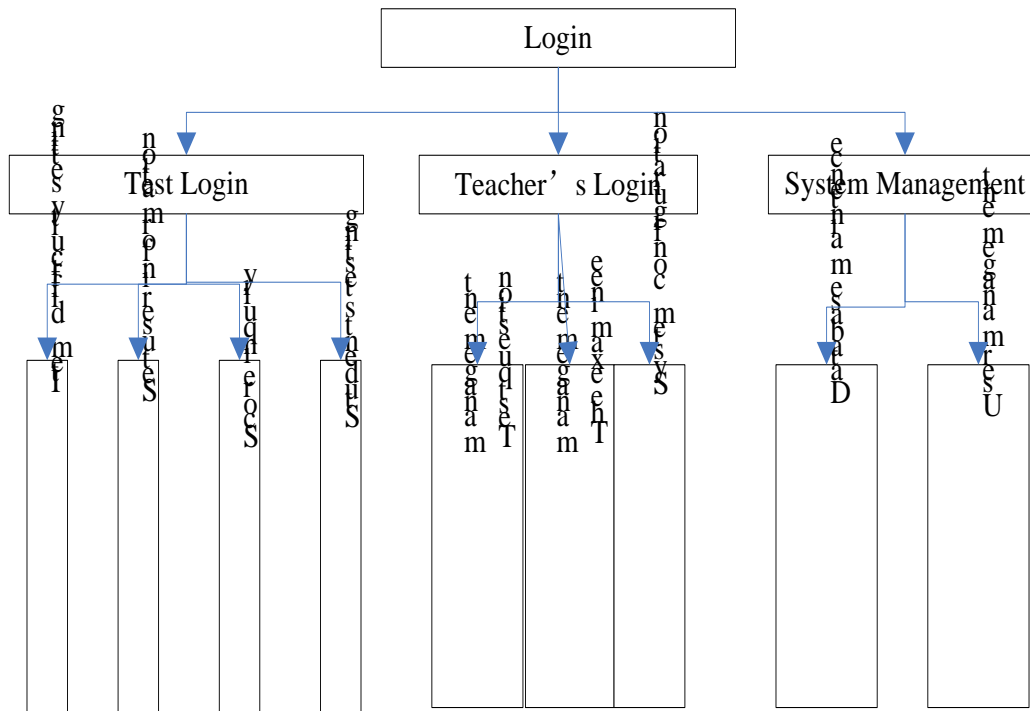


Figure 1: System structure graphic

The following function is mainly realized by system test module:

1). Random extraction of test items

For the random extraction of test items, it is realized through the storage in SQL. Through the random rearrangement of items tables in database, forming a random test items sequence table, and then extract top 50 or more items from it.

2) Topic display

Due to the dynamic generation of test items, this system adopts dynamic control, and then binds the topic dynamic state into the options, so as to lay a good foundation for the further calculation of result.

Take a selection item as an example in details, firstly, place a panel **control** onto the interface, so that the controlling piece related to items information is placed into this controlling piece. With Label control for topic information and RadioButtonList control for options, extract content a, b, c and d from the item bank, for the dynamic items, add it into ArrayList, and then bind ArrayList onto RadioButtonList to realize the dynamic display of test items. Function code segment is as follows:

```

Label lbl = new Label();
    lbl.Text = "NO." + i.ToString()+ dr["TRY_SUBJECT"].ToString().Trim();
    lbl.ID = "lbl" + i.ToString();
RadioButtonList rb = new RadioButtonList();
    rb.ID = i.ToString();
    rb.Attributes["runat"] = "server";
    rb.Items.Add("A. " + dr["a"]);
    rb.Items.Add("B. " + dr["b"]);
    rb.Items.Add("C. " + dr["c"]);
    rb.Items.Add("D. " + dr["d"]);
    for (int j = 1; j < 5; j++)
    { rb.Items[j - 1].Value = j.ToString(); }
    Pan.Controls.Add(lbl);
    Pan.Controls.Add(rb);
    ans.Add(dr["TRY_KEY"]);
    cent.Add(dr["TRY_CENT"]);
  
```

3). Display and record the result

How to display and record the result? Seek out the user's opinions among the test items, and then compare it with standard answers in the item bank, keep the score if it is correct. For the specific realization, with the adoption of FindControl through RadioButtonList name, seek out SelectedValue to look for the answers selected by the user. After the completion of grading, insert it into a grade table according to user name, topic type and value.

Test function module based on IRT

The overall test system mainly includes two parts, namely foreground and background, among which, the foreground provides the user with various kinds of adaptive test and result display; the background mainly includes item bank, management and maintenance and statistical analysis, etc.

In the process of computer adaptive test, the following problems must be solved by the system:

①How to start, namely, what kind of strategy is followed to select the first item; ②How to continue it, how to select next item after the examinee completes one item; ③How to determine the termination condition for the ending of test. Aiming at above problems, for this system, it intends to adopt the following test scheme:

1). Experimental exploration stage

For the initial item selection strategy adopted by this system: if the examinee participates in the test, then determine the examinee's initial competence value according to the historical record, so as to determine the item started for this time; if the examinee has not participated in the test, then the examinee determines examinee's competence degree voluntarily to select the initial test item.

After the determination of initial difficulty, this system will take 10 items as a group of test item firstly, after 10 items are completed by the examinee, if answers given by the examinee are all right or wrong, then select one extremely difficult or easy item, so that this stage can be terminated actually. If all answers given by the examinee are still right or wrong after reselection of 10 items, then it shows that the test items in item bank are too difficult or too easy for the examinee, his/her actual level can't be measured, and then the test is terminated at this time, with feedback of test result to the examinee.

If the answers given by the examinee are partially right and partially wrong, then estimate its peculiarity level initially by use of maximum likelihood method, and end experimental exploration stage. Function code segment is as follows:

```
if (!IsPostBack)
{
    byte[] u = new byte[100];
    Session["u"] = u;
    Session["F1"] = 0F;
    Session["F2"] = 0F;
    this.ViewState["topic"] = 0;
    if (Session["paramGo"] == null)
        Session["paramGo"] = 0.8F;
    Session["ability"] = 0.5F;
    Session["score"] = 0;
}
IRTFetch((float)Session["paramGo"]);
```

2). Subsequent topic selection strategy

After the initial estimation of examinee's competence level, what is the method of selection of next test item for the purpose of accurate estimation of examinee's competence level?

Item information function is an index which reflects the accuracy of estimation of examinee's competence level directly by virtue of test score, it has direct ratio relations with the examinee's competence level estimated, the larger the item information function is, the more accurate the estimation on examinee's competence level will be. Under the condition of known item quality parameters, the information content provided by one item is not unchanged actually, and its value

depends on the examinee's peculiarity level. That is to say, for the same item, when testing the examinees with high level, medium level and low level, the information content provided is different. Under the condition of unchanged item quality parameters, the value taken by item information function is varied along with examinees' different peculiarity level, the closer the examinees' peculiarity level is to the difficulty of test item, the larger the value taken by item information function will be.

Therefore, if the item bank used to measure this peculiarity level has been established, and some peculiarity level of one examinee is known, then it can select a series of items with the largest information content provided for the examinee.

After the determination of test item selection principle, when doing one item, the system will estimate the examinee's competence by use of maximum likelihood method according to the examinee's reaction, continue to select the test item, and add the information content, until the test termination condition is met.

3) Test termination strategy

The test degree is decided by our requirements on measurement accuracy. As the cumulative sum of topic information content is the test information function value, the estimated standard error is equal to reciprocal of square root of information function. Along with the increasing of test topic, the cumulative value of information becomes larger constantly, when this cumulative value reaches the specified requirements, the test can be terminated. For the insufficiency of this method, it reflects in too long test caused, particularly for the examinee with extremely high or extremely low competence level, the efficiency is relatively low.

The test process based on IRT is shown in Figure 2:

Summary

Adaptive test system is a new kind of test theory, along with the development of item response theory and artificial intelligence technology, etc., it will be developed largely. For adaptive test system, it will be developed towards the more intelligent direction. In addition to the application into formal test, it shall be also taken as simulation test system for usual practice, according to each student's personal learning situation, apply an intelligent expert system to realize specific training environment aiming at individual, so as to make its application wider.

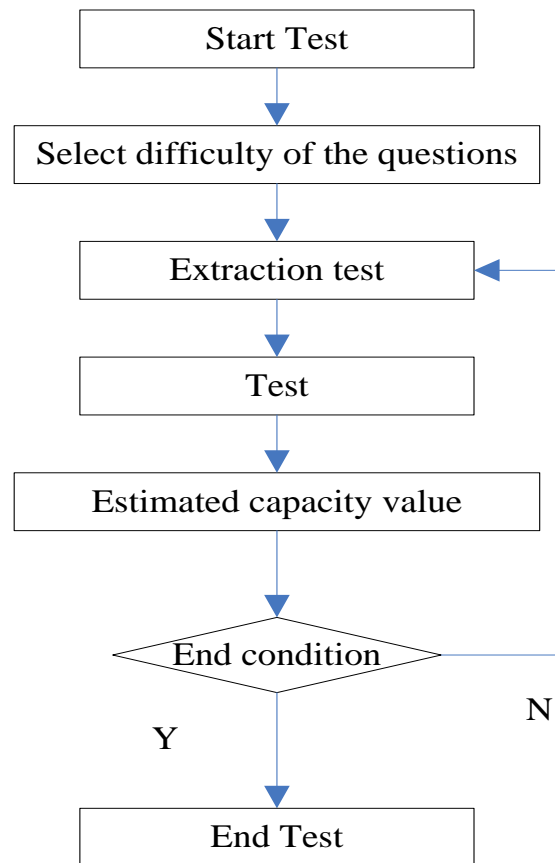


Figure 2: Test flow chart based on IRT

References

- [1] FENG Y. *Application of Item Response Theory into the Building of Item Bank System*[J]. Journal of Jiangsu Institute of Commerce, 2014.2.
- [2] Michael Kolen and Robert Brennan *Test equating: methods and practices* (Springer, 1995) and *Item response theory: Parameter estimation techniques* by Frank Baker, Marcel Dekker, 1992.
- [3] ZHAO Z.Y. & ZHANG Y.F.. *Computer Application Test Based on WEB Technology*[J]. Digital system users, 2013.12.
- [4] ZHANG Y.P. et al. *Dynamic Website Built by ASP.NET+SQL*[M]. Beijing: Beijing Electron Industry Press, 2006.
- [5] XING C.Y.. *Design and Realization of Computer Foundation Test System Based on WEB*[D]. Hebei University of Science and Technology, 2012.12.