Reducing Item Exposure in Computerized Adaptive Testing Systems Using Automatic Item Generation

Harutyun Terteryan

Department of Computer Systems and Informatics, State Engineering University of Armenia, Yerevan 1109, Armenia

Abstract: This paper studies the technics of reducing item exposure by utilizing automatic item generation methods. Known test item calibration method uses item parameter estimation with the statistical data, collected during examinees prior testing. Disadvantage of the mentioned item calibration method is the item exposure; when test items become familiar to the examinees. To reduce the item exposure, automatic item generation method is used, where item models are being constructed based on already calibrated test items without losing already estimated item parameters. A technic of item model extraction method from the already calibrated and therefore exposed test items described, which can be used by the test item development specialists to integrate automatic item generation principles with the existing testing applications.

Key words: Item exposure, calibration, IRT (item response theory), adaptive testing, automatic item generation, item model.

1. Introduction

One of the inconveniences of CAT (computerized adaptive testing) is that it makes tremendous demands on item count in test item bank. There are several models that adaptive testing systems are based on [1]. In this paper, the author refers to one of them called IRT (item response theory) [2].

IRT is generally claimed as an improvement over CTT (classical test theory). For tasks that can be accomplished using CTT, IRT mainly brings greater flexibility and provides more sophisticated information. Computerized adaptive testing is enabled by IRT and cannot reasonably be performed using only classical test theory [3]. Another advantage of IRT over CTT is that the more sophisticated information IRT provides allows a researcher to improve the reliability of an assessment. IRT entails three assumptions:

- A unidimensional trait denoted by θ_j ;

- Local independence of items;

- The response of a person to an item can be modeled by a mathematical IRF (item response function):

$$P(\theta_j) = c_i + \frac{(1-c_i)}{1+e^{1.7a_i(b_i-\theta_j)}} \quad (1)$$

where:

- b_i —item difficulty (item location) parameter;
- a_i — item discrimination parameter;
- c_i — item guessing parameter;
- θ_j — examinees ability parameter.

There are three types of IRT logistic models depending on the number of parameters used in the model. The 3PL (Fig. 1) is named so because it employs three item parameters. The two-parameter model (2PL) assumes that the data have no guessing, but that items can vary in terms of location (b_i) and discrimination (a_i). The one-parameter model (1PL) assumes that guessing is a part of the ability and that all items that fit the model have equivalent discriminations, so that items are only described by a single parameter (b_i). Section 2 discusses the item exposure issue CAT.

Corresponding author: Harutyun Terteryan, graduate student, research fields: knowledge evaluation, item response theory, e-testing systems, computerized adaptive testing. E-mail: harroot@gmail.com.

Fig. 1 3 PL model IRF.

Section 3 introduces the automated Item generation model. Section 4 describes different approaches of integrating AIG (automatic item generation) with e-testing applications. Section 5 introduces the suggested item model extraction tool. Section 6 gives conclusions.

2. Item Exposure

The procedure of evaluating parameters of IRF is called item calibration. Test item bank calibration procedure requires presence of examinees answers collected during prior test [4]. So a group of examinees need to pass the test, which makes them familiar to the test items included in the item bank thus increases the item exposure quota.

In order to reduce item exposure during the prior test, the exposed items should be replaced with the new ones. It will farther increase the required count of test items in the item bank. Also taking in account that item creating process is time consuming, it will degrade productivity of test development specialists. Replacing test items in the item bank is the easiest way to solve the item exposure issue, but in case of IRT it will bring another drawback. By replacing test items, we will lose the item parameter values estimated during the calibration process.

To enhance test item creation efficiency, AIG methods are being used. Automatic item generation requires two steps [5]. First, test development specialists create item models. Second, these item model elements are manipulated to generate new items with the aid of computer-based algorithms. With this

two-step process, a dozen of new items can be created from a single item model.

3. AIG Item Modeling

An item model (a.m. item schema/template) serves as an explicit representation of the variables in an assessment task, which includes the stem, the options, and oftentimes auxiliary information. The stem is the part of an item that formulates context, content, and/or the question the examinee is required to answer. The options contain the alternative answers with one correct option and one or more incorrect options or distracters [6].

When dealing with a multiple-choice item model, both stem and options are required. With an open-ended or constructed-response item model, only the stem is created. Auxiliary information includes any additional material, in either the stem or option, required to generate an item, including digital media such as text, images, tables, diagrams, sound, and/or video. The stem and options can be divided further into elements, also known as item model variables. By systematically manipulating the elements, measurement test item development specialists can generate large numbers of items from one item model.

If the generated items or instances of the item model are intended to measure content at similar difficulty levels, then the generated items are isomorphic. To illustrate a simple example of isomorphic test item model, Fig. 2 has been given (see Fig. 2).

There are several ways to create a relationship between options and stem parameters. One of them is to provide the formula like in Fig. 2, another is to have an options list and keep a link to the array of stem parameter values. The advantage of AIG is also that it gives opportunity to generate new test items on the fly, during adaptive testing process [7].

4. Integrate Existing Application with AIG

If the testing system is initially designed to use AIG, the workflow of the application is like on the diagram shown in Fig. 3. First, the test development specialists

The lengths of the right triangle sides are $a = 12$ and $b = 10$, respectively.
What is the length of the hypotenuse?
A. 222
B. 242
C. 244
D. 262

Stem

The lengths of the right triangle sides are $a = E1$ and $b = E2$, respectively.
What is the length of the hypotenuse?

Elements

E1 parameter value(s): [4, 400] by step 5.

E2 parameter value(s): [3, 300] by step 10.

Options

A. $(E1)^2 + (E2)^2 - (E1 + E2)$

B. $(E1)^2 + (E2)^2 - |E1 - E2|$

C. $(E1)^2 + (E2)^2$

D. $(E1)^2 + (E2)^2 + 18$

Key(s)

C

Fig. 2 Simple item model. **Fig. 3 AIG applied application workflow.** **Fig. 4 AIG integrated application workflow.**

need to create item models, it must be noted that only the isomorphic models are being considered in this workflow. Other models, which items might have different IRF parameters, are using the extended model of so-called LLTM (linear logistic test model) known as random-effects LLTM or LLTM-R [6].

Fig. 3 shows the workflow of the testing application that already have the AIG mechanisms applied. But what if the application that is not using AIG also wants to implement adaptive testing?

Nowadays there are many testing applications that are

not adaptive, not using AIG but, have a huge bank of test items (already exposed) and collected a statistical data about them. It can be a very valuable source for adaptive test item calibration. Now if they want to utilize the advantages of CAT and AIG, there must be a technic to extract already created test items to the AIG item models. So the application workflow should be something like on the diagram shown in Fig. 4.

The suggested technic is opposite to the usual model creation process, if the usual method assumes that test model development specialists have already created item models and only then generate the actual test item instances, the suggested technic assumed that there is already calibrated test item and an isomorphic item model needs to be extracted from it.

5. Item Model Extraction

Item model extraction procedure is meant to define the model variables described in Fig. 2. It is very similar to the usual model creation procedure. As a result of the current research, a new MET (model extraction tool) has been created that can be used by item model development specialists to extract the model from the test item. Fig. 5 shows the UI snippet of MET.

Fig. 5 Model extraction tool.

Fig. 5 shows the 3rd and 4th steps of the Fig. 4 diagram. On the 3rd step the test development specialist selects the test item for further processing. MET provides a user interface to highlight the corresponding parts of the stem and extract them as model variables.

6. Conclusions

In order to reduce item exposure quota in testing systems that have intent to utilize adaptive item selection methods, a new technic has been suggested. A generic AIG item model extraction method has been advised to not lose the statistical information for the already existing test items and generate new item instances for further adaptive testing purposes. As a result of the current research, a new tool has been created that gives an opportunity to the test item development specialists to create AIG isomorphic item models and generate new test item instances from them in order to reduce the proportion of item exposure in the system's test item bank. The results of this research and the model extraction tool are being used in student's knowledge evaluation application in State Engineering University of Armenia during migration of conventional e-testing application to the computerized adaptive testing tool. The new model extraction tool makes it much easier for the test

development specialists to convert the existing conventional test items to the adaptive test items, without losing the already calibrated IRF parameters.

References

- [1] Economides, A. A., and Roupas, C. 2007. "Evaluation of Computer Adaptive Testing Systems." *International Journal of Web-Based Learning and Teaching Technologies* 2 (1): 12-5.
- [2] Baker, F. B. 2001. *The Basics of Item Response Theory*. Portsmouth, NH: Heinemann, 5-12.
- [3] Hambleton, R. K. and Jones, R.W. 1993. "Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development." *Educational Measurement: Issues and Practice* 12 (3): 38-47.
- [4] Makransky, G. 2009. "An Automatic Online Calibration Design in Adaptive Testing." In *Proceedings of the Conference on Computerized Adaptive Testing (GMAC)*. 2009.
- [5] Bejar, I. I. 2002. "Generative Testing: From Conception to Implementation." In *Item Generation for Test Development*, edited by Irvine, S. H., and Kyllonen, P. C. Hillsdale, NJ: Erlbaum, 199-217.
- [6] Gierl, M. J., and Lai, H. 2012. "The Role of Item Models in Automatic Item Generation." *International Journal of Testing* 12 (3): 273-98.
- [7] Bejar, I. I., Lawless, R., Morley, M. E., Wagner, M. E., Bennett, R. E., and Revuelta, J. 2003. "A Feasibility Study of On-The-Fly Item Generation in Adaptive Testing." *Journal of Technology, Learning and Assessment* 2 (3): 1-29. <http://www.jtla.org>.