T. Hirashima et al. (Eds.) (2011). Proceedings of the 19th International Conference on Computers in Education. Chiang Mai, Thailand: Asia-Pacific Society for Computers in Education

Detecting Redundant Items in Construction of Multiple Equivalent Test Forms using Latent Dirichlet Allocation

Pokpong SONGMUANG[a*], Maomi UENO[b] & Keizo NAGAOKA[a]
[a]*School of Human Sciences, Waseda University, Japan*
[b]*Graduate School of Information Systems, The University of Electro-Communications, Japan*
*pokpong@aoni.waseda.jp

Abstract: We propose an automatic construction method of multiple equivalent test forms indicated by test information function, and the method reduces the probabilities of selecting redundant items to the same test form. In previous studies, although their methods minimized the different between the test information functions of the constructed test forms, they neglected the content similarities of selected items in the same test form. Therefore, the content similar items have probabilities to be selected into the same test form. This affects the test reliability. The main idea of this paper is to reduce the probabilities by applying a latent Dirichlet allocation in the test construction method to detect the content similarities between the selected items and the remaining items in the item banks.

Keywords: e-testing, multiple test forms, test construction, latent Dirichlet allocation

Introduction

Educational assessments sometimes need "multiple test forms" in which each form consists of a different set of items but still has qualities that are equivalent (e.g., equivalent amounts of test information based on item response theory) to the others. For example, multiple test forms are needed when a testing organization administers a test in different time slots. To achieve this, multiple test forms are constructed in which all forms have equivalent qualities so that examinees, who have taken different test forms, can be objectively evaluated on the same scale.

In order to construct multiple test forms, e-testing, which accomplishes automated test construction, has recently become popular in research areas involving educational measurement (Luecht, 1998; van der Linden and Adema, 1998; van der Linden, 2005). The methods in previous studies have been used to construct all forms of a test to satisfy the same test constraints (e.g., the number of test items and the amount of test information) to ensure that all forms have equivalent qualities. van der Linden and Boekkooi-Timminga (1989) proposed a sequential method of constructing test forms using linear programming to minimize the fitting errors to the test constraints. While, Boekkooi-Timminga (1990) and Armstrong (1992) proposed methods that simultaneously constructed all test forms to minimize the differences in the fitting errors on the test forms. The former used linear programming and the latter used network-flow programming. van der Linden and Adema (1998) and van der Linden (2005) proposed methods that sequentially constructed test forms by minimizing the difference in fitting errors between a currently constructed test form and the remaining set of items in the item bank.

Although the reviewed methods minimized the differences between the fitting errors on the test forms, these methods neglected considering the content similarities of the selected items in the same form. Therefore, in actual constructions of multiple test forms, the content similar items have probabilities to be selected into the same test form. When the content similar items are selected to the same test form, they are redundant and might become hints for the other item(s). This reduces the reliability of the test form.

The main idea of this paper is to solve the problem by applying natural language processing technique to detect the content similarities of items while constructing test forms. Recently, in natural language processing, one of the famous statistical models, which can be applied to detect the content similarities of items, is the latent Dirichlet allocation (LDA) (Blei, et al., 2003). Several researches (Blei, et al., 2003; Kakkonen, et al., 2006) showed LDA provided better performance than the previous statistical models such as Latent Semantic Analysis (LSA) (Deerwester, 1990) and probabilistic Latent Semantic Analysis (pLSA) (Hofmann, 1999).

However, it is well known that the multiple test construction has the trade-off problem between the equivalent of test forms and computational costs. To alleviate the trade-off problem, Songmuang and Ueno (2010) proposed a Bees algorithm (BA) for multiple test construction that applies a parallel-computing technique that distributes the computational costs to multiple processors without increasing the differences in fitting errors.

Therefore, to reduce probabilities of selecting redundant items into the same test form and construct multiple equivalent test forms in a realistic time, we combine LDA with BA for multiple test construction. This method constructs multiple test forms to minimize the probabilities of selecting the content similar items in the same test form and to minimize the different of the test information function between test forms.

Construction Method of Multiple Equivalent Test Forms based on Bees Algorithm combining Latent Dirichlet Allocation

1. Bees algorithm for multiple test forms construction

In this section, we describe a method of constructing test forms based on BA that constructs multiple equivalent test forms by minimizing the difference in fitting errors between test forms. The construction of multiple test forms is classified as an NP-hard problem. To reduce the computational time, we divided the construction of test forms into two steps:

Step A: Construct test forms only to minimize the fitting errors of each form to test constraints without taking into consideration the equivalence of test forms. Here, the constructed test forms are still not equivalent.

Step B: Extract the most equivalent set of test forms from the constructed test forms in Step A that minimizes the difference in fitting errors between test forms.

The proposed construction of multiple test forms based on BA is implemented in a parallel-computing environment that includes one server and several workers.

Using a parallel-computing technique, the computational cost of constructing the test forms for each processor core is calculated by dividing the computational cost by the number of processor cores. Therefore, we can decrease the computational time by increasing the total number of processor cores of workers. As a result, we can relax the trade-off by using the proposed method and the parallel computing technique.

However, this method neglects the content similarities of the selected items in the same test form. Next, section we describe how to apply LDA to detect the content similarities in BA.

2. Latent Dirichlet Allocation

LDA is a generative probabilistic model in natural language processing which is used for evaluate topic of data. The basic idea of this model is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words.

Several researches compared the effectiveness of LDA with previous models such as LSA and pLSA (Deerwester, 1990; Hofmann, 1999). According to the results, LDA provides better performance to estimate topic for data.

Therefore, in this paper, we apply LDA to evaluate topics of items. The evaluated topics of items are used to indicate the content similarities of items.

To apply LDA in multiple test constructions, we combine LDA in step A of BA to minimize the probabilities of selecting content similar items of each test form. In this method, we generate the selection probabilities of items inversely proportional to the content similarities of the selected items in the test form and the remaining items in an item bank. That is the remaining items, which have different topics comparing to the topics of selected items, have high probabilities to be selected into the test form.

Finally, we develop the construction system of multiple test forms using the proposed method and perform a simulation experiment using an actual item bank to show the effectiveness of the proposed method.

References

[1] Armstrong, R. D., Jones, D., & Wu, I.-L. (1992). An automated test development of parallel tests from a seed test, *Psychometrika*, *57*(2), 271–288.

[2] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

[3] Boekkooi-Timminga, E. (1990). The construction of parallel tests from irt-based item banks, *Journal of Educational Statistics*, *15*(2), 129–145.

[4] Deerwester, S., Dumais, S., Landauer, T., Furnas, G. & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society of Information Science*, *41*(6), 391–407.

[5] Hofmann, T., Probabilistic Latent Semantic Indexing. *Proceedings of the 22nd International Conference on Research and Development in Information Retrieval* (SIGIR'99), 1999.

[6] Kakkonen, T., Myller, N. & Sutinen, E. (2006). Applying Latent Dirichlet Allocation to Automatic Essay Grading. *Advances in Natural Language Processing*, Lecture Notes in Computer Science, *4139/2006*, 110-120.

[7] Luecht, R. (1998). Computer-assisted test assembly using optimization heuristics. *Applied Psychological Measurement*, *12*(3), 224–236.

[8] Songmuang, P. & Ueno (2010). M., Bees Algorithm for Construction of Multiple Test Forms in e-Testing, *IEEE Transactions on Learning Technologies*. (Accepted)

[9] van der Linden, W. J. (2005). *Linear Models for Optimal Test Design*, 1st ed., Statistics for Social and Behavioral Sciences. Springer.

[10] van der Linden, W. J. & Adema, J. J. (1998). Simultaneous assembly of multiple test forms. *Journal of Educational Measurement*, *35*(3), 185–198.

[11] van der Linden, W. J. & Boekkooi-Timminga (1989). E., A maximin model for test design with practical constraints, *Psychometrika*, *54*, 237–247.