Proc. Natl. Sci. Counc. ROC(D)
Vol. 11, No. 2, 2001. pp. 78-87

A Greedy Approach to Test Construction Problems

KOUN-TEM SUN

Institute of Computer Science and Information Education
National Tainan Teachers College
Tainan, Taiwan, R.O.C.

(Received March 21, 2000; Accepted November 16, 2000)

Abstract

The *item response theory* (IRT) is a well-developed theorem, which can be used to estimate the unidimensional ability of a student precisely. Based on this theory, the construction of parallel test forms can be performed such that the teaching activities can be evaluated precisely. Unfortunately this test construction problem is a combinatorial optimization problem, and now there is no polynominal time algorithm that exists for finding the optimal solution (i.e., minimizing the deviations from the target (parallel) test information function). In this paper, an effective method is proposed for constructing parallel test forms with approximation to the amount of target test information by selecting items from an item bank. The experimental results show that the proposed method can be used to obtain very good results that approximate those obtained using recently proposed approaches. However, our method is much simpler than the other methods, thus, our proposed method significantly reduces the computation time for constructing the parallel test forms. This method will be very useful to educational measurement.

Key Words: item response theory, test construction, combinatorial optimization problem, educational measurement

I. Introduction

The *item response theory* (IRT) (Hambleton & Swaminathan, 1985; Lord, 1952; Lord & Novick, 1968; Weiss, 1982) is a well-developed theorem, which can be used to precisely estimate the unidimensional ability of a student. This theory has proved to be very useful in educational assessment. For example, when there is a close fit between the chosen IRT model and the test data set of interest, the following merits can be obtained:
(1) Item parameter estimates are independent of the group of examinees.
(2) Examinee ability estimates are independent of the particular choice of test items.
(3) Precision of ability estimates is known (predictable).

Owing to these advantages, IRT has been widely used in educational assessment (Macro, 1977; Wright & Douglas, 1977; Wright & Stone, 1979; Yen, 1981), especially in constructing a desired test for measuring the ability of a student on the basis of the item information (Birnbaum, 1968; Lord, 1980). The larger the value of the item information, the more precise the measurement of ability. By applying the item information function, conventional methods could be used to construct a desired test efficiently. The constructed test information function is the sum of the item information function $I_j(\theta)$ for the items comprising the test (Birnbaum, 1968):

$$I(\theta) = \sum_{j=1}^{m} I_j(\theta), \tag{1}$$

where m is the number of items in the constructed test.

For example, if a scholarship test for an academic award were to be constructed, then the items with greater information at high ability levels would be selected (Fig. 1) in order to screen students with high abilities. For another case, if an educational research project requires two groups of students with abilities at the middle-high and the middle-low, then the two-peak (i.e., a “M” shape) target test information function would be required.

Lord (1977) outlined a procedure, originally conceptualized by Birnbaum (1968) for the use of the item information function in the test building process. The steps are:

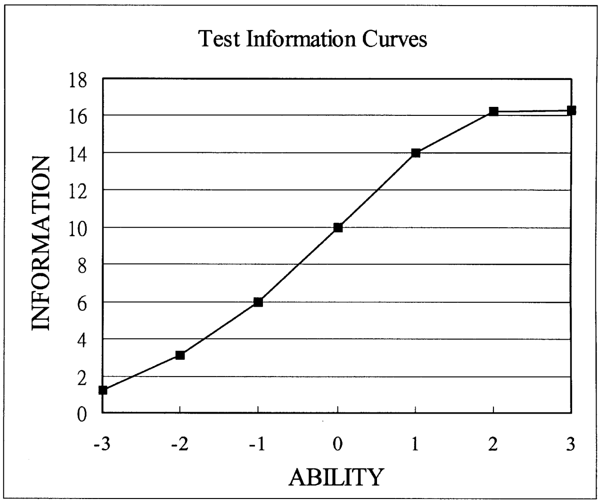

Fig. 1. The curve of information function for a scholarship test.

1. Define the shape of the desired test information function (or called target information function).
2. Select items with item information functions that will fill up the hard-to-fill areas under the target information function.
3. After each item is added to the test, calculate the test information function for all the selected test items.
4. Continue selecting test items until the test information function approximates the target information to a satisfactory degree.

The fewer the errors between the target test information function and the constructed test information function, the greater the satisfaction of the test. Therefore, a test designer would select items such that the constructed test information function approached the target test information function. When the item bank is small, manual and sequential item selection and examination of the test information curve against the target are not difficult. When the item bank is large, however, it is very difficult to accomplish this job without the aid of some heuristic methods. Since the item selection problem is a combinatory optimization problem and the number of combinations increases exponentially with the size of problem, therefore, only the weak methods (heuristic algorithms) are used for finding the "good" solutions. For example, the linear programming (LP) is the technique used most often during the test construction (Van der Linden & Boekkooi-Timminga, 1989; Baker, Cohen & Barmish, 1988; Boekkooi-Timminga, 1987; Swanson & Stocking, 1993; Wang & Ackerman, 1997). Based on the linear programming, items are selected into a test to achieve the optimization objectives under the problem's constraints. Some heuristic methods have also be used to find good solutions, for example, the branch-and-bound method (Adema, 1989), the revised simplex method (used the relaxed 0-1 linear programming model (Adema, 1990)), and the weighted deviation model (Swanson & Stocking, 1993). Test construction problems commonly involve a list of objective functions with various purposes (Van der Linden & Boekkooi-Timminga, 1989), but the test information function is the common objective of all test design problems. Therefore, in this paper, we only considered how to select items in order to meet the requirements of the desired test information function. However, the difficulty of this problem was not reduced by eliminating some considerations such as the exposure rate and the content attributes. For constructing a test information function approximating the target test information function, we propose an effective method based on the concept of energy function of neural network models (Hopfield & Tank, 1985; Lippmann, 1987; Sun & Fu, 1993a; Sun & Fu, 1993b). Energy function approaches have been used successfully to solve many optimization problems (Papadimitriou & Steiglitz, 1982; Sun, 1992; Sun & Fu, 1992; Sun & Fu, 1993a; Sun & Fu, 1993b; Wu, Xia, Li, & Chen, 1996). However, we transformed the test construction problem into an energy function and then solved it using a greedy approach. The results obtained using our approach were similar to those obtained using recently proposed methods. However, the computation time of our proposed method was much shorter than that of their methods. In Section 2, some recently proposed methods on test construction problems are described. In Section 3, we explain the concept and the detailed operations of the proposed method. In Section 4, experimental results obtained using our proposed approach and other methods are included. Finally, a conclusion is given in Section 5.

II. Test Construction Methods

There have been many methods proposed for automating item selection in test construction (e.g., Ackerman, 1989; Adema, 1990; Boekkooi-Timminga, 1989; Theunissen, 1985; Van der Linden, 1987; Yen, 1981; Swanson & Stocking, 1993; Wang & Ackerman, 1997; Sun & Chen, 1999). The most commonly used methods for item selection are based on the concept of binary programming. This principle offers many advantages for solving the test construction problems using classical linear programming algorithms. Some methods may apply heuristic techniques (e.g., simulated annealing (Jeng & Shih, 1997), replacement algorithm, neural network (Sun & Chen, 1999), etc.) to obtain fewer errors when gathering information between the constructed test and the target test. Recently, three methods have been proposed that could construct parallel test forms efficiently and would be explained. To simplify the test construction process, we supposed that all items in the bank contained the same attributes. Then, the goal of test construction problem for parallel test forms was to minimize the deviations (errors) between the target test information function and the constructed test information function. Based on this assumption, previously proposed methods for constructing parallel tests were simplified as follows.

A. Swanson and Stocking Method (Swanson & Stocking, 1993)

There are two phases, selection phase and replacement phase, included in this method.

1. The **selection phase** consists of the following four steps:
   (1) For every item t not yet included in the test, the expected error that will be generated as if the item t were selected into the test. Then, the expected error at the latent trait (ability) j is represented by the value of q_{tj}:

$$q_{ij} = \sum_{i=1}^{n} a_{ij}x_i + (m-k)v_j + a_{tj}, \quad 1 \le t \le n, \qquad \textbf{(2)}$$

   where n is the number of items in the item bank, a_{ij} is the information quantity of item i at the latent trait (ability) level j, x_i denotes the status of item i whether item i is included in ($x_i = 1$) or excluded from ($x_i = 0$) the test, m is the number of items required for the test, k is number of selected items in the test, and v_j is the average item information for all items in the item bank at the specified ability level j.
   (2) Compute the difference across all ability levels j between the target test information function and the constructed test information function if the item t were added to the test.

$$D_{tj} = Abs\ (d_j - q_{tj}),\ 1 \le t \le n, \qquad \textbf{(3)}$$

   where $Abs()$ is the absolute value function, and d_j is value of the target test information function at the ability level j.
   (3) Select the item t^* with the smallest value of D_{tj}, and add it to the test.

$$D_{t^*j} = Min\{D_{tj},\ 1 \le t \le n\}. \qquad \textbf{(4)}$$

   (4) Repeat Steps 1 to 3 until m items have been selected.
2. The **replacement phase** consists of the following three steps:
   (5) Select the $(m + 1)^{th}$ item according to Steps 1 to 3, except eliminate the term "$(m - k)v_j$" in Equation (2), and then add this additional item into the test.
   (6) Find an item already included in the test whose removal will most reduce the error between the target test information function and the constructed test information function.
   (7) If the removal and replacement process reduced the error between the target test information function and the constructed test information function, then add the replacement item (in Step 5) to the test, delete the removal item (in Step 6), and repeat Steps 5 and 6. Otherwise, stop the process (no more errors can be reduced using the replacement process).

The replacement phase will monotonically improve the error between the target test information function and the constructed test information function. If it is not possible to find a pair of items whose replacement in the test would result in a smaller error, then the process stops.

B. Wang and Ackerman Method (Wang & Ackerman, 1997)

Wang and Ackerman proposed two algorithms: TESTGEN1 and TESTGEN2. The TESTGEN2 is similar to TESTGEN1 except when considering multiple content information functions. Under our assumption "items contain the same attributes", The TESTGEN1 and TESTGEN2 would be the same, thus, only the TESTGEN1 is introduced in this paper. The Wang and Ackerman method consists of the following six steps:

1. Randomly select the first item into the test.
2. Compute the test information function o_j.

$$o_j = \sum_{i=1}^{n} a_{ij}x_i\ . \qquad \textbf{(5)}$$

3. Check the number of items included in the test. If the number of items is equal to m, then stop. Otherwise, go to Step 4.
4. Compute the difference, E_j, between the target test information function and the constructed test information function, and then prioritize the order of items according to the size of difference at each ability level j.

$$E_j = Abs(d_j - o_j) \quad \textbf{(6)}$$

5. Select the item t at the ability level j^* where the greatest priority occurs. The item t could minimize the difference E_{j^*} if it were added to the test.

$E_{j^*} = Max\{E_j, j = 1 \sim s\}$, where s is the number of ability levels. **(7)**

$$T_{ij^*} = Abs(E_{j^*} - a_{ij^*}),\ \forall x_i \neq 1, \text{ and } i = 1 \sim n. \quad \textbf{(8)}$$

$$T_{tj^*} = Min\{T_{ij^*}\ ,\ i = 1 \sim n\}. \quad \textbf{(9)}$$

6. Add the item t into the test, and go to Step 2.

Based on the TESTGEN1 algorithm, during each iteration only the item with the most information will be selected into the test, thus, the greatest difference of the information quantities between the target test and the constructed test are reduced.

C. Neural Network Method (Sun & Chen, 1999)

An artificial neural network technique is proposed and used to solve the test construction problem. First, the test construction problem is transformed into an energy function, and then the energy function is reduced using the neural network technique. When the energy function reaches a stable state, the state of the neurons represents the solution to the problem. This method can be used to obtain very good results, while the computation is more complex. The detailed steps of this method are stated as follows.

1. Set the initial values of all variables in the energy function.

Variables d_j, $j = 1 \sim s$, are set to the desired information of the test, and x_i, $i = 1 \sim n$, are set to zero (i.e., no items had been selected into the test from the item bank). All w_{ij}, $i = 1 \sim n$ and $j = 1 \sim s$, are equal to the information quantity of the item i at the ability level j. The initial time t is set to zero.

2. Determine the value of the energy function E_I at time t.

$$E_I(t) = \sum_{j=1}^{s} (d_j - O_j(t))^2\ , \quad \textbf{(10)}$$

where $O_j(t)$ is constructed test information function at time t.

3. Determine the updated value of each x_i (i.e., Δx_i) for reducing the energy of function $E_I(t)$,

$$\Delta x_i(t) = -\frac{\dfrac{\partial E_I(t)}{\partial x_i(t)}}{\dfrac{\partial^2 E_I(t)}{\partial x_i^2(t)}} = \Big(\sum_{j=1}^{s} (d_j - O_j(t)) w_{ij}\Big) \Big/ \Big(\sum_{j=1}^{s} w_{ij}^2\Big)\,. \quad \textbf{(11)}$$

4. Select the most "appropriate" neuron (variable x_i) to update.

Find a neuron $x_{i^*}(t)$with the greatest absolute value of $\partial E_I(t)/\partial x_i(t)$, for all i, and the value $x_{i^*}(t)$ is not equal to one, and the updated value of $x_{i^*}(t)$ (i.e., $\Delta x_{i^*}(t)$) is greater than or equal to one.

$$\frac{\partial E_I(t)}{\partial x_{i^*}(t)} = Max.\left\{\left|\frac{\partial E_I(t)}{\partial x_i(t)}\right|,\ \text{for } \forall x_i \neq 1 \text{ and } \Delta x_i \geq 1\right\}, \quad \textbf{(12)}$$

5. Update the state of neuron $x_{i^*}(t)$.

The new value of $x_{i^*}(t)$ is set to one (i.e., the item i^* is then selected into the test), and becomes

$$x_{i^*}(t+1) = 1. \quad \textbf{(13)}$$

6. Compute the difference of the energy function E_I between two consecutive iterations.

$$\Delta E_I(t+1) = E_I(t+1) - E_I(t). \quad \textbf{(14)}$$

7. Check the stability of the energy function.

If the following equation is true, then the energy of the function $E_I(t)$ is increasing but decreasing. The updating iteration will stop.

$$\Delta E_I(t+1) \geq 0. \quad \textbf{(15)}$$

When it is true, go to Step 8. Otherwise, $t = t + 1$ and go to Step 2 for the next iteration.

8. Stop.

All variables x_i, with the value "1" are the items selected into the constructed test.

This neural network approach represents the test construction problem using an energy function that considers the total difference between the target test information function and the constructed test information function. After applying the converging process, it can be used to find a good solution.

These proposed methods can be used to efficiently solve test construction problems, and find the good solutions. However, the computing processes are complex. In the following section, an effective method termed the "greedy approach" was designed to obtain good results at a much faster pace.

Table 1. An Example of the Target Test Information Function (two peaks)

	Index of Ability Level (i)				
	1	2	3	4	5
Ability Level (θ_i)	–2.0	–1.0	0.0	1.0	2.0
Test Information ($d(\theta_i)$)	4	12	6	12	4

III. Test Construction by the Greedy Approach

The proposed greedy approach is similar to the neural network method. The errors between the target test information function and the constructed test information function are also considered and represented by an energy function. When the target test information function (d(θ_i), as shown in Table 1) is specified by the test designer, the information quantity of the target test d_i (the simplified representation of $d(\theta_i)$) can be obtained.

After a test is constructed, the constructed test information function $O(\theta_i)$ can also be determined. Then, the sum of squared error of information between the target test information function and the constructed test information function can be calculated using $E_I = \sum_{j=1}^{s}(d_j - O_j)^2$ (as shown in Equation (10)).

The value of O_j, the constructed test information function, can be derived from Equation (16).

$$O_j = \sum_{\iota=1}^{n} w_{ij}x_i\,, \quad x_i \in \{0,\ 1\} \tag{16}$$

where n is the total number of items in the item pool, and w_{ij} is the information of item i at the ability-level index j. When an item i is selected into the test (i.e., the status of x_i is changed from 0 (excluded from the test) to 1 (included in the test)), the energy of function E_I, the sum of squared error of information between the target test information function and the constructed test information function, would be changed (either increased or decreased) by a value:

$$\Delta E_{I,i} = E_{I,i} - E_I, \tag{17}$$

where

$$E_{I,i} = \sum_{j=1}^{s}(d_j - O_j(t))^2\,, \tag{18}$$

and

$$O_{j,i} = \sum_{\substack{k \neq i \\ k=1}}^{n} w_{kj}x_k + w_{ij}\,. \tag{19}$$

Then, the updated value of energy function E_I after selecting the item i is

$$\begin{aligned}\Delta E_{I,i} &= \sum_{j=1}^{s}(d_j - O_{j,i})^2 - \sum_{j=1}^{s}(d_j - O_j)^2 \\ &= \sum_{j=1}^{s}(d_j^2 - 2d_jO_{j,i} + O_{j,i}^2 - d_j^2 - 2d_jO_j - O_j^2) \\ &= \sum_{j=1}^{s}(-2d_jw_{ij} + 2O_jw_{ij} + w_{ij}^2) \\ &= \sum_{j=1}^{s} w_{ij}(w_{ij} + 2O_j - 2d_j)\end{aligned} \tag{20}$$

If the test information function after the item i added were less than the desired test information function, then the updated value of energy function E_I (i.e., $\Delta E_{I,i}$) would be less than zero. Under this condition, if the item i were selected into the test, then the energy of the function E_I would decrease until the value of $\Delta E_{I,i}$ was greater than or equal to zero. Based on this concept, the greedy approach was designed to select items with the greatest information to fill the gap between the constructed test information function and the target test information function. This method can be used to efficiently construct a test approximating the target test information function.

A. The Greedy Approach to Item Selection

In the proposed greedy approach, each variable x_i (i.e., the state of item) is computed independently such that the updated value of $E_{I,i}$, $1 \le 1 \le n$, can be determined in parallel. However, not all variables can be updated simultaneously. Otherwise, the energy function E_I would be over-updated and then oscillate. It would not converge to a stable state. So, only one variable x_i is updated at each iteration, and the energy function E_I is reduced monotonically until it reaches a minimum state. The detailed operations of the proposed greedy approach are stated as follows.

1. Set the initial values of all variables.

Variables d_j, $j = 1 \sim s$, are set to the target test information function specified by the test designer, and x_i, $i = 1 \sim n$, are set to zero initially (i.e., initially, no item is selected into the test from the item bank). The value of w_{ij}, $i = 1\sim n$ and $j = 1 \sim s$, is equal to the amount of information of the item i at ability level j. The initial time t (iteration index) is set to zero.

2. Determine the constructed test information function $O_j(t)$ at the iteration t.

$$O_j(t) = \sum_{i=1}^{n} w_{ij}x_i(t)\,, \quad \forall j = 1 \sim s. \tag{21}$$

Table 2. Item Parameters, Information, and Classical Statistics of Items (the first 20 items of the 320 items in the natural science item bank)

Item	Item Parameters			Ability Level							Classical Statistics	
	a	b	c	-3	-2	-1	0	1	2	3	p	r
1	0.763	−3.000	0.270	0.242	0.193	0.078	0.024	0.007	0.002	0.001	0.969	0.584
2	0.435	−0.546	0.200	0.024	0.052	0.083	0.093	0.075	0.049	0.027	0.656	0.331
3	0.803	0.537	0.310	0.000	0.003	0.035	0.171	0.250	0.130	0.041	0.575	0.381
4	1.187	0.080	0.350	0.000	0.002	0.059	0.466	0.293	0.052	0.007	0.684	0.534
5	0.844	−1.176	0.180	0.031	0.191	0.364	0.215	0.068	0.017	0.004	0.816	0.681
6	0.730	0.169	0.290	0.001	0.012	0.075	0.199	0.192	0.090	0.030	0.628	0.444
7	0.732	1.127	0.370	0.000	0.001	0.010	0.062	0.169	0.164	0.076	0.531	0.243
8	0.625	−1.650	0.260	0.062	0.147	0.164	0.101	0.044	0.017	0.006	0.859	0.514
9	1.428	2.831	0.370	0.000	0.000	0.000	0.000	0.001	0.102	0.715	0.384	0.016
10	0.541	0.063	0.280	0.006	0.024	0.069	0.117	0.114	0.072	0.035	0.634	0.347
11	0.983	−1.587	0.310	0.035	0.265	0.342	0.116	0.025	0.005	0.001	0.906	0.694
12	0.661	−1.707	0.250	0.070	0.171	0.182	0.102	0.041	0.014	0.005	0.869	0.529
13	0.538	−1.368	0.230	0.048	0.105	0.134	0.105	0.058	0.027	0.011	0.803	0.436
14	1.183	−0.378	0.290	0.000	0.012	0.247	0.549	0.156	0.024	0.003	0.744	0.669
15	0.400	−0.363	0.350	0.012	0.027	0.047	0.058	0.054	0.039	0.024	0.719	0.150
16	0.558	0.220	0.270	0.004	0.020	0.064	0.122	0.127	0.082	0.040	0.603	0.365
17	0.960	0.378	0.280	0.000	0.003	0.045	0.287	0.340	0.116	0.026	0.575	0.481
18	0.814	1.828	0.300	0.000	0.000	0.002	0.019	0.126	0.267	0.174	0.384	0.230
19	0.891	−0.490	0.310	0.002	0.033	0.205	0.301	0.132	0.035	0.008	0.753	0.570
20	1.083	−1.295	0.240	0.016	0.231	0.526	0.195	0.036	0.006	0.001	0.866	0.739

3. Determine the value of the energy function E_I at the iteration t.
   As shown in Equation (10), we can compute the value of energy function $E_I(t)$.

$$E_I(t) = \sum_{j=1}^{s} (d_j - O_j(t))^2 . \quad \textbf{(22)}$$

4. Determine the updated value of the energy function E_I, (E_I, $\Delta E_{I,i}$, after selecting the item i, ($\forall i$, $i = 1 \sim n$.

$$\Delta E_{I,i}(t) = \sum_{j=1}^{s} w_{ij}(w_{ij} + 2O_j(t) - 2d_j) \quad \textbf{(23)}$$

5. Find the smallest value $\Delta E_{I,i^*}(t)$.
   Find the smallest value $\Delta E_{I,i^*}(t)$ from Equation (23) for all i where $x_i(t)$ is not equal to one (i.e., item i excluded from the test).

$$\Delta E_{I,i^*}(t) = min\{\Delta E_{I,i}(t),\ x_i(t) \neq 1,\ i = 1 \sim n\}. \quad \textbf{(24)}$$

6. Check the stability of the energy function.
   Two cases need to be considered:
   (1) The number of test items is not fixed (but limited by an upper bound m).
   The energy function is checked to see whether the energy function has reached a minimum state or not. If the following equation is true, then the energy function $E_I(t)$ is not decreasing but increasing after the item i^* was added.

$$\Delta E_{I,i^*}(t) \geq 0. \quad \textbf{(25)}$$

   The iterative operations for reducing the energy of the function $E_I(t)$ will stop. Otherwise, the item selection operations continue until a minimum state is reached or the number of test item reaches the upper bound m.
   (2) The number of test items is fixed.
   Under this condition, we only need to check the following equation to see whether it is true or not.

$$(\sum_{i=1}^{n} x_i) - m = 0 , \quad \textbf{(26)}$$

   where m is the number of items required for the test. When either Equation (25) or Equation (26) is true, go to Step 7. Otherwise, update the state of variableinto $x_{i^*}(t)$ one (i.e., item i^* is then selected into the test), $t = t + 1$, and go to Step 2 for the next iteration.
7. Stop.
   All variables x_i with the value one would be the items included in the test.

By following the steps of the proposed greedy approach, only one item is selected into the test during each iteration. Thus, the maximum number of iterations is equal to m for constructing a test with m items,

Table 3. The Ranges of the Desired Test Information Function with One Peak

	Index of Ability Level				
	1	2	3	4	5
Ability Level	−2.0	−1.0	0.0	1.0	2.0
Test Information	4 ~ 5	6 ~ 8	18 ~ 21	6 ~ 8	4 ~ 5

Table 4. The Ranges of the Desired Test Information Function with Two Peaks

	Index of Ability Level				
	1	2	3	4	5
Ability Level	−2.0	−1.0	0.0	1.0	2.0
Test Information	5 ~ 6	11 ~ 13	7 ~ 9	11 ~ 13	5 ~ 6

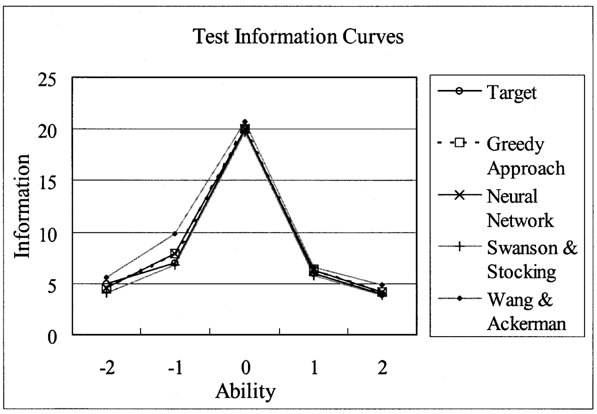

Fig. 2. Test information curves (one peak) produced with the greedy approach and other methods under the condition of fixed number (= 40) test items.

and the magnitude of computations (Horowitz & Sahni, 1978) in each iteration is O(n) for computing n operations in functions $O_j(t)$, $j = 1 \sim s$ (s is a constant), $\Delta E_{I,i}(t)$, and $\Delta E_{I,i^*}(t)$. The total computation complexity of the proposed greedy method is O(mn) (= $m \times$ O(n)) which is the same as other methods (Sun & Chen, 1999). However, the computation time of our method is much less than other methods and the results are as good as those obtained using them.

IV. Performance Evaluation

The items, used to evaluate the performance of our method and other methods, were randomly selected from the item pool as part of the project "The Web-based Natural Science Learning Environment in Elementary School" supported by the National Science Council of Taiwan, ROC, under the grant of NSC 89-2520-S-024-001-. The item discrimination parameter (a_i) ranged from 0.40 to 2.50, the item difficulty parameter (b_i) ranged from -3.0 to 3.0, and the pseudo-chance parameter (c_i) ranged from 0.08 to 0.44 (see Table 2). The amount of information of the desired tests varied from the ranges defined in Table 3 (for one-peak information curve) and Table 4 (for two-peak information curve). Based on the limitations of information quantities defined in these two tables, 100 target test information functions were randomly generated, respectively.

Figs. 2, 3, 4 and 5 show the simulation results using the greedy approach and other methods under the conditions of fixed and not fixed number of items, respectively. The averages of sum of squared errors between the target test information functions and the constructed test information functions are shown in

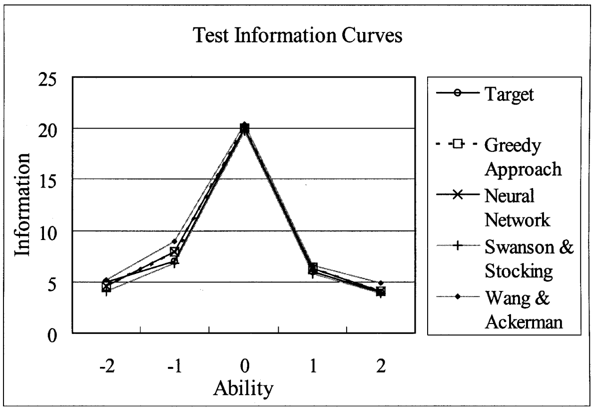

Fig. 3. Test information curves (one peak) produced with the greedy approach and other methods under the condition of not fixed number test items (but limited by an upper bound 40).

Table 5, which can be computed using Equation (27):

$$Average\_Error = \frac{1}{N}\sum_{k=1}^{N}\sum_{j}^{s}(d_{kj} - O_{kj})^2 \qquad \textbf{(27)}$$

where N is the number of test samples, and s is the number of ability levels. Thus, d_{kj} is the desired test information function for sample k at the ability level j, and O_{kj} is the constructed test information function for sample k at the ability level j. The average value of errors made using our method were much fewer than those made using Wang and Ackerman's method, and approached the value made using the neural network method and Swanson & Stocking's method. We found that fewer errors were used using the proposed greedy approach than those made using the Wang & Ackerman (1997) method. In addition, they were approximately

Table 5. The Averages of Sum of Squared Errors between the Desired Test Information Functions and the Constructed Test Information Functions with the Ranges Defined in Table 3 and Table 4. The Computation Time is the Average of Execution Time for Constructing One Test on the Computer (simulated on Pentium II PC)

	Greedy Approach	Neural Network	Swanson & Stocking		Wang & Ackerman
			Before Replacement	After replacement	
Fixed (Table 3)	0.6986	0.6945	1.1772	0.9715	12.8340
Not fixed (Table 3)	0.6841	0.6879	1.1772	0.9715	6.8865
Fixed (Table 4)	0.7251	0.7416	0.4840	0.2755	2.9536
Not foxed (Table4)	0.7099	0.7327	0.4840	0.2755	0.7000
Average Errors	0.7044	0.7142	0.8386	0.6235	5.8435
Computation Time (sec.)	0.41	0.91	0.53	1.23	
[speed up}*	[1.00]	[0.45]	[0.77]	[0.33]	

(*[speed up] = computation time of greedy approach / other method)

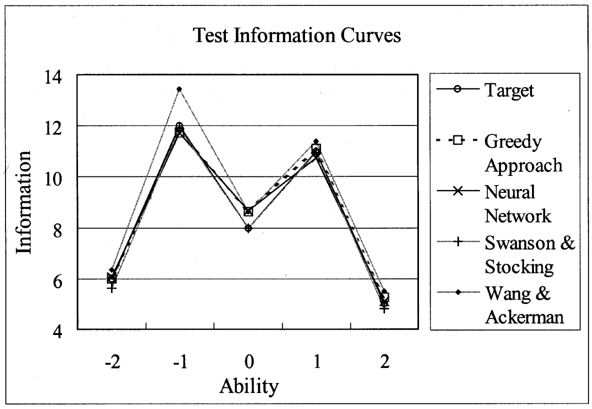

Fig. 4. Test information curves (two peaks) produced with the greedy approach and other methods under the condition of fixed number (= 40) test items.

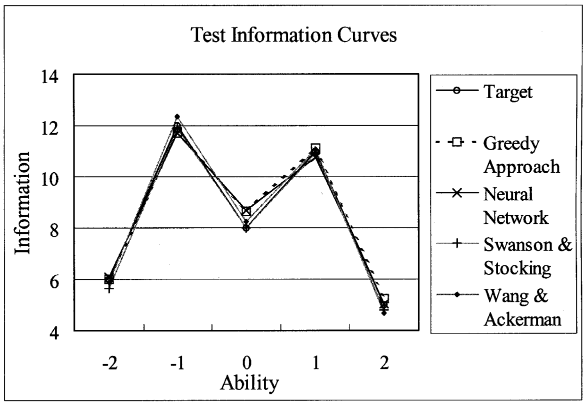

Fig. 5. Test information curves (two peaks) produced with the greedy approach and other methods under the condition of not fixed number test items (but limited by an upper bound 40).

the same as those made using the neural network approach (Sun & Chen, 1999) and the Swanson & Stocking's method, respectively. The computation time of our proposed approach was much less than those of other methods. The improvement ratio of computation time is at least 23% faster. Thus, the proposed item selection method was more effective than the other methods, and will be very useful to the test designers to construct the parallel test forms or a desired test for educational assessment.

V. Conclusions

In this paper, the greedy approach, based on the IRT information function, was proposed to construct a test from an item bank. The proposed method can be effectively used to construct parallel test forms or desired tests such that the constructed test information function approximates the target test information function. A real item pool was used to evaluate the performance of our method and other methods. The experimental results showed that the errors of our proposed approach were very small and approximated those obtained using other methods. However, the computation time of our method was much less than theirs. Our proposed method significantly shortened the computation time of test construction, while maintaining the quality (errors between test information functions) of parallel test forms. Thus, our greedy approach will be a very useful tool for researchers with test construction problems.

Acknowledgments

This research is supported by the National Science Council of Taiwan, ROC, under the grant of NSC 89-2520-S-024-001-.

References

Ackerman, T. (1989). *An alternative methodology for creating parallel test forms using the IRT information function.* The annual meeting of the National Council for measurement in Education, San Francisco.

Adema, J. J. (1989). *Implementations of the branch-and-bound method for test construction problems*. Research Report 89-6. Enschede: Department of Education, University of Twente, The Netherlands.

Adema, J. J. (1990). *A revised simplex method for test construction problems*. Research Report 90-5. Enschede: Department of Education, University of Twente, The Netherlands.

Baker, F. B., Chen, A. S., & Barmish, B. R. (1988). *Item characteristics of tests constructed by linear programming. Applied Psychological Measurement*, **12**, 189-199.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of metal test scores* (pp. 395-479). Reading MA: Addison-Wesley.

Boekkooi-Timminga, E. (1987). Simultaneous test construction by zero-one programming. *Methodika*, **1**, 101-112.

Boekkooi-Timminga, E. (1989). *Models for computerized test construction*. The Haag, The Netherlands: Academisch Boeken Centrum.

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory— Principles and Applications*. Kluwer Academic Publishers Group, Netherlands.

Hopfield, J. J., & Tank, D. W. (1985). Neural computation of decision in optimization problems. Biological Cybernetics, **52**, 141-152.

Horowitz, E., & Sahni, S. (1978). *Fundamentals of Computer Algorithms, Pseudocode*. Computer Science Press, Inc.

Jeng, H. L., & Shih, S. G. (1997). A comparison of pair-wise and group selections of items using simulated annealing in automated construction of parallel tests. *Psychological Testing*, **44** (2), 195-210.

Lippmann, R. P. (1987). An introduction to computing with neural nets. *IEEE ASSP magazine*, **4**, 4-22.

Lord, F. M. (1952). A theory of test scores, *Psychometric Monograph*, **7**.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading Mass: Addison-Weslet.

Lord, F. M. (1977). Practical applications of item characteristic curve theory. *Journal of Educational Measurement*, **14**, 117-138.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

Macro, G. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, **14**, 139-160.

Papadimitriou, C. H., & Steiglitz, K. (1982). *Combinatorial Optimization: Algorithms and Complexity*. Prentice-Hall, Inc., NJ: Englewood Cliffs.

Sun, K. T. (1992). *A Study of Optimization Problems by Neural Networks*. Unpublished doctoral dissertation, National Chiao Tung University, Taiwan, ROC.

Sun, K. T., & Fu, H. C. (1992). A Neural Network Implementation for Traffic Control Problem on Crossbar Switch Networks. *International Journal Neural Systems*, **3**(2), 209-218.

Sun, K. T., & Fu, H. C. (1993a). A Hybrid Neural Network Model for Solving Optimization Problems. *IEEE Transactions on Computers*, **42**(2), 218-227.

Sun, K. T., & Fu, H. C. (1993b). A Neural Network Approach to the Traffic Control Problem on Reverse Baseline Networks. *Circuits, Systems and Signal Processing*, **12**(2), 247-261.

Sun, K. T., & Chen, S. F. (1999). A study of applying the artificial intelligent technique to select test items. *Psychological Testing*, **46**(1), 75-88.

Swanson, L., & Stocking, M. L. (1993). A model and heuristic for solving very large item selection problems. *Applied Psychological Measurement*, **17**(2), 151-166.

Theunissen, T. J. J. M. (1985). Binary programming and test design. *Psychometrika*, **50**, 411-420.

Van der Linden, W. J. (1987). Automated test construction using minimax programming. In w. J. Van der Linden (Ed.), *IRT-based test construction* (pp. 1-16). Enschede, The Netherlands: Department of Education, University of Twente.

Van der Linden, W. J., & Boekkooi-Timminga, E. (1989). A maximum model for test design with practical constraints. *Psychometrika*, **54**, 237-247.

Wang, C. S., & Ackerman, T. (1997). Two item selection algorithms for creating weakly parallel test forms using the IRT information functions. *Psychological Testing*, **44**(2), 123-140.

Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement*, **6**, 379-396.

Wright, B. D., & Douglas, G. A. (1977). Best procedures for sample-free item analysis. *Applied Psychological Measurement*, **1**, 281-295.

Wright, B. D., & Stone, M. (1979). H. *Best test design*. Chicago: MESA.

Wu, X. Y., Xia, Y. S., Li, J., & Chen, W. K. (1996). A high-performance neural network for solving linear & quadratic programming problems. *IEEE Trans. on Neural Networks*, **7**(3), 643-651.

Yen, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement*, **5**, 245-262.

A Greedy Approach to Test Construction

利用貪婪技術於測驗建構問題之研究

孫光天

台南師範學院資訊教育研究所

摘　要

題目反應理論（Item Response Theory；IRT）爲一發展完善之測驗理論，能用來精確評估一位受試者之潛在能力，因此，若能建構兩份訊息量相同（相近）之平行測驗，將可有效評估教學活動之成效；但是，以題目訊息量爲基礎之測驗建構問題爲一組合最佳化問題，至今仍無一多項式（polynominal）運算時間之演算法可求得最佳解（使平行測驗間之測驗訊息量誤差最少）。在本文中，我們將提出一以貪婪技術爲基礎的測驗建構方法，經擬實驗結果顯示，我們所提的方法，能產生訊息量誤差非常小的平行測驗，不亞於其他新近所提的方法，但我們方法所需運算時間卻遠少於其他方法，能快速求得答案，肯定了本研究所提之貪婪技術，可有效的建構測驗，對教育評量技術之進展將非常有助益。