A Heuristic Of CAT Item Selection Procedure For Testlets

Yuehmei Chien
David Shin
Walter Denny Way

Introduction

Testlet

Definition: Items are bundled together due to the related context such as a common stimulus, passage, theme, scenario, or content category.

Introduction (cont'd)

Administering testlets in CAT

- In terms of the size of testlets

  For smaller size testlets, the whole set of items are administered to the examinee at once.

  For larger size testlets:

  - subsets of items associated with the testlet are *pre-determined*
  - subsets of items associated with the testelet are *assembled on the fly*

- In terms of the adaptiveness
  - adaptive between testlets
  - adaptive between and within testlets

Introduction (cont'd)

Content balancing in CAT

Simple method:

constrained CAT (CCAT; Kingsbury & Zara, 1989):a straightforward method looking for the content category for which the cumulative percentages of administered items currently is farthest below its target percentage.

More elaborate methods:

- the a-parameter stratification strategy with content blocking (a_STR_C; Yi & Chang, 2003; Cheng, Chang, & Yi, 2007)
- the Weighted Penalty Model (WPM; Shin, Chien, Way & Swanson, 2009)
- the Weighted Deviations Model (WDM; Stocking & Swanson, 1993)
- the Shadow test approach (van der Linden, 2000)

Introduction (cont'd)

the Shadow Test Approach

- two important advantages:
  - meet all constraints
  - maintain measurement precision

- Definitions: A shadow test is a complete test form assembled in real time prior to each item selection that

  - satisfies all constraints;
  - has maximum information at the current ability estimate; and
  - contains all items previously administered to the examinee.

Introduction (cont'd)

Implementation of the shadow test approach

- uses the integer solver (van der Linden, 2005a, 2005b; Veldkamp & van der Linden, 2008) in the commercial software package CPLEX 9.0 (ILOG, Inc., 2003)
- the average time for selecting an item is about 0.43 seconds for one shadow test and 1.01 seconds for two shadow tests (Veldkamp & van der Linden, 2008)

Introduction (cont'd)

Purpose

Propose a heuristic that performs item selection for testlets with the following features:

- testlet level constraints can be specified for each testlet besides the test level constraints for the whole test;
- shadow tests are assembled in real time for testlet selection, which means testlet is selected from the shadow test; hence all test level and testlet level constraints are met; and
- once a testlet is selected, the item selection is adaptive within that testlet

Implementation of the Proposed Heuristic

Challenge:

the integer solver is not used, how to assemble the shadow tests in real time?

Strategies:

- Shadow test assembling is a typical combinatorial optimization problem that searches from a very large complete space of solutions.
- Branch and Bound is by far the most widely used tool for the combinatorial optimization problem, which is a general algorithm that consists of a systematic enumeration of all possible solutions, where large subsets of solutions impossibly optimal are discarded.
- The solutions possibly optimal are discovered first.

  Sort subsets of testlets by average information and use the subsets with larger information to assemble the shadow tests.

the Heuristic—Preparation

testlet level content constraints

A testlet is like a mini pool and a subset of a testlet is like a mini test form. The constraints can be imposed into the mini test form so the subsets associated with the same testlet administered to the examinee can be similar across different examinees. For instance, six items are selected from a 20-items passage-based testlet measuring the four reading aspects—forming a general understanding, developing interpretation, making reader/text connections, and examining content and structure. Constraints: one item measuring the first aspect and one to two items measuring the second, the third, and the fourth aspects, respectively.

the Heuristic—Preparation

Assembling subsets for each testlet

Subsets are assembled:

- for each theta level
- satisfying the testlet level content constraints
- not violate the upper bounds of the test level constraints

The maximum number of subsets and the minimum subset information are both specified in order to limit the number of candidate subsets in a reasonable size and discard the subsets that are impossible to be optimal.

the Heuristic—Item Selection Procedure

Two phases

- the testlet selection
- the subset and item selection

After the testlet is selected, this selected testlet is set active. Given this active testlet and the current estimated theta, one of the optimal subsets is chosen and then the most informative item is selected from that chosen subset.

the Heuristic—Testlet Selection

For each testlet selection, the following information is calculated or given:

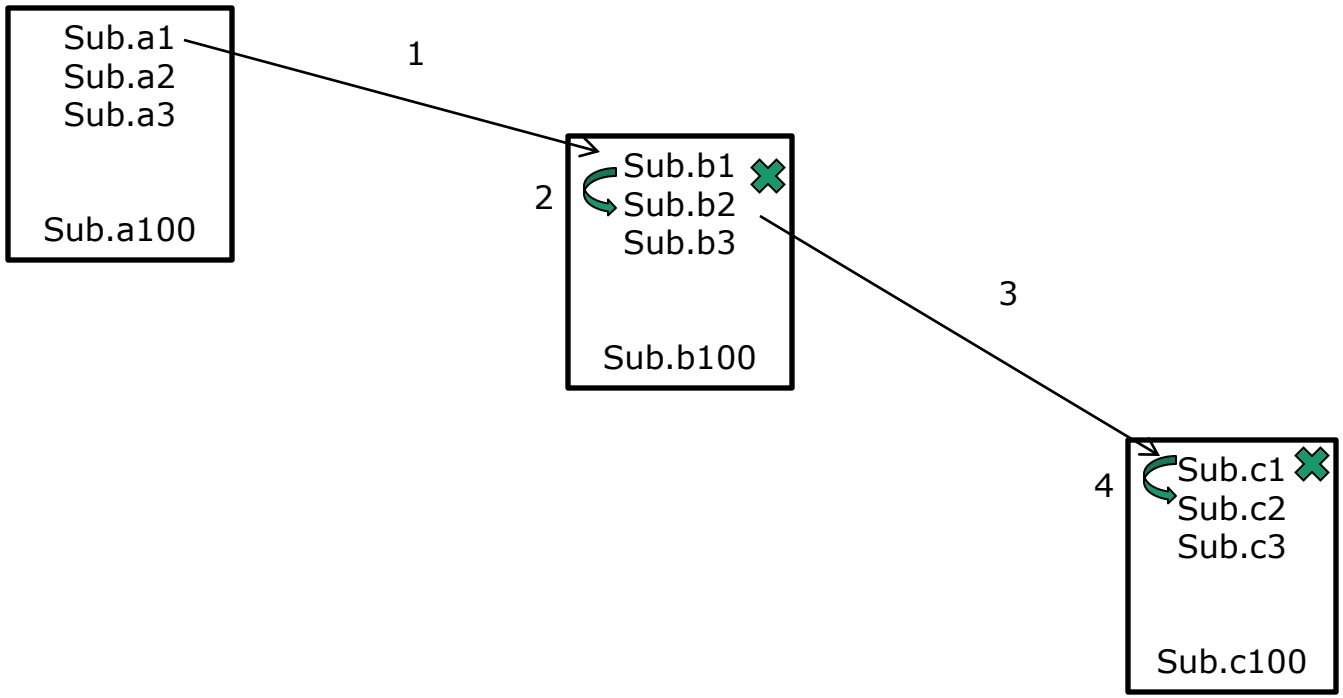
- *current bounds: for each* test level constraint, upper and lower bounds minus the total number of items associated with each test level constraint based on the previously administered items (Any negative number is set to 0.)
- for the remaining test, the number of testlets and the number of items from each testlet that should be administered.
- the current ability estimate and its corresponding theta level

the Heuristic—Testlet Selection (cont'd)

1. Assembling a shadow test composing the remaining test
2. Giving the current theta level, get the top *A* informative subsets across different testlets, where A is a reasonable large number such as 200.
3. Sequentially pick one subset at a time.
   a) If it is not the last testlet selection and this picked subset is not violating the upper bounds of the current bounds, assume this subset is selected and administered. Go to the next testlet selection with the updated current bound.
   b) If it is the last testlet selection and this picked subset does not satisfy the current bounds, pick the next subset. If all subsets have exhausted, go back one testlet level to the previously subset selection to get the next subset.
   c) If it is the last testlet selection and this picked subset satisfies the current bounds, a shadow test is found.

the Heuristic—Testlet Selection (cont'd)

An Example of forming a 3-testlet shadow test



the Heuristic—Subset and Item Selection (cont'd)

Input: given the shadow test of the remaining test that composes *B* subsets from *B* different testlets

1. set the testlet associated with the first subset active
2. reselect a subset from this active testlet
   - calculate the allowed upper bounds and lower bounds for selecting the subset given the previously administered items and the subsets in the shadow test except for the first subset
   - obtain the subsets associated with the current theta level
   - those subsets are sorted by the average information given the currently estimated theta
   - exclude the subsets that do not meet the allowed upper bounds and lower bounds
   - select one subset from several subsets at the top
   - pick one item from the chosen subset

   For the rest of item selection after the first item is selected, the same steps above are applied but the eligible subsets are those that include the items previously administered from this active testlet.

Results

Pool: 40 testlets, 804 items

Test: 30 items composing 6 subsets

Testlet level constraints:

6 testlet level constraints and each item in the testlet associated with one testlet level constraint

Test level constraints:

0.1	**0.3**	**VC**
0.04	**0.15**	**RP**
0.04	**0.10**	**RC.01**
0.06	**0.2**	**RC.02**
0.10	**0.25**	**RC.03**
0.10	**0.35**	**RC.04**
0.06	**0.3**	**LA**
0	**0.2**	**Functional**
0.2	**0.4**	**Informational**
0.6	**0.8**	**Literary**
0.15	**0.5**	**key_a**
0.15	**0.5**	**key_b**
0.15	**0.5**	**key_c**
0.15	**0.5**	**key_d**

Results (cont'd)

	Not Constrained	Constrained
bias	0.0328	-0.0005
MSE	0.1472	0.1792
Correlation	0.9343	0.9258
CSEM	0.3871	0.3917

Averagely, one item is selected around 0.008 to 0.04 second.

Questions?