

複数等質テストを用いた適応型テストのシミュレーション評価 Simulation Evaluation of Adaptive Testing using Multiple Test Forms

宮澤 芳光^{*1}, 石井 隆稔^{*2}, 植野 真臣^{*3}
Yoshimitsu MIYAZAWA^{*1}, Takatoshi ISHII^{*2}, Maomi UENO^{*3}
^{*1}東京学芸大学
^{*1}Tokyo Gakugei University
^{*2}東京理科大学
^{*2}Tokyo University of Science
^{*3}電気通信大学
^{*3}University of Electro-Communications
Email: miyazawa@u-gakugei.ac.jp

あらまし:近年, Web 上でテストを実施する e テスティングにおいて適応型テストが注目されている。適応型テストとは, 受検者の解答履歴から逐次的に能力を推定し, その能力に適した項目をアイテムバンクから自動的に出題することにより, より少ない項目で精度の高い能力推定を行う手法である。この適応型テストでは, 能力が同等な受検者には全く同じ項目群が出題されるため, アイテムバンクにある項目が有効に活用されていなかった。アイテムバンクにある項目は非常に高価であり, この項目を有効に活用することが重要になる。そこで, 本研究では, アイテムバンクにある項目を有効活用するため, 複数等質適応型テストの構成手法を提案する。具体的には, まず, 異なる項目により構成されているにもかかわらず, 各テストが等質であるような複数等質テストを構成する。次に, その複数等質テストをアイテムバンクとして適応型テストを行う手法を提案する。本論文ではシミュレーション実験により, 提案システムの有効性を示す。

キーワード: e-テスティング, 適応型テスト, 項目反応理論, 教育測定

1. はじめに

近年, Web 上でテストを実施する e テスティングの実用化が進み注目されている⁽¹⁾。e テスティングにおいて, 項目はアイテムバンクと呼ばれるデータベースに蓄積されている。e テスティングでは, このアイテムバンクから受検者の能力に応じて項目を抽出する適応型テスト (Computerized adaptive testing; CAT) と呼ばれるものが特に注目されている。適応型テストとは, 受検者の項目への解答履歴から逐次的に能力を推定し, その能力に適した項目をアイテムバンクから選択するテスト出題戦略である。適応テストの利点は, 受検者の能力推定精度を減少させずにテストに費やされる時間や項目数を減らすことなどが挙げられている。しかし, この適応型テストでは, 能力が同等な受検者には全く同じ項目群が出題されるため, アイテムバンクにある項目が有効活用されていなかった。アイテムバンクにある項目は非常に高価であり, この項目を有効に活用することが重要になる。

そこで, 本研究では, 能力が同等な受検者であっても異なる項目を出題する複数等質適応型テストの構成手法を提案する。具体的には, まず, 異なる項目により構成されているにもかかわらず, 各テストが等質であるような複数等質テストを構成する。ここでは, 受検者の能力測定精度が等質になるようにテスト情報量を基準に複数等質テストを構成する。次に, その複数等質テストをアイテムバンクとして適応型テストを行う手法を提案する。本手法では, 構成された複数等質テストの数の等質型適応型テ

ストを作ることができるため, 等質型適応型テストの構成数は複数等質テストの数に依存する。複数等質テストを構成する手法は, これまでに研究されてきている。本研究では, 石井らの手法⁽³⁾を用いる。石井らの手法⁽³⁾は, 従来の手法と比較し, アイテムバンクから構成できる等質テストの数が最大である事が保障されているため, この手法を用いる事により数多くの等質適応型テストを作る事ができる。さらに, 石井らの手法⁽³⁾は, e テスティングでのテストに使われる手法であるため, 受検者の解答数は, 最長でも構成されたテストの項目数である。そのため, 受検者の能力測定精度の下限は石井らの手法⁽³⁾によって構成されたテストのテスト情報量である。そのため, 本研究のアイデアは, 石井らの手法⁽³⁾を用いて構成されたテストを適応型テストの枠組みを用いて受検者に出題する項目数を減らすというものである。

本論では, シミュレーション実験により, 本システムの有効性を示す。

2. 等質適応型テスト

アイテムバンクを有効活用するため, 複数等質適応型テストの出題手法を提案する。以下に本手法のアルゴリズムを示す。

- (1) 石井らの手法⁽³⁾を用いてアイテムバンクから複数等質テストを構成
- (2) 構成された複数等質テストをアイテムバンクとして情報量が最大の項目を選択
- (3) 受検者は選択された項目を解答

- (4) 受検者の解答が正誤判定され、それまでの解答履歴を用いて受検者の能力を推定
 (5) 受検者の能力の推定値が収束しているならば終了、そうでなければ(2)へ
 上記のアルゴリズムを用いて受検者の能力を推定する。

3. シミュレーション評価

本章では、提案手法と従来の適応型テストを比較し、提案手法の有効性をシミュレーションによって示す。また、本手法では、複数等質適応型テストの構成数が複数等質テストの構成数に依存するため、複数等質テストの構成手法に Linden らの手法⁽⁴⁾を用いて比較する。

複数等質テストの構成条件は以下の通りである。

- アイテムバンクの項目数：2000
- 複数等質テスト間の重複項目数：5
- 複数等質テストの項目数：25
- 情報量の上限・下限：表1
- 重複項目数：5, 0

重複項目数とは、複数等質テスト間で共通に使うことができる項目の数である。適応型テストは、表1の情報量の下限に達したとき、テストを終了する。ここでは、シミュレーションにより適応型テストが1000回実施されたときの出題パターンとアイテムバンク中の項目の使用回数を分析する。

実験の結果を表2に示す。実験の結果、重複項目数を5とした石井らの手法⁽³⁾を用いた提案手法は、出題パターンが1000パターンであった。これは、全ての受検者が異なる出題パターンを解答したことを意味する。また、この提案手法は、項目の使用回数が最も高く、アイテムバンクの項目を有効に活用している。Linden らの手法⁽⁴⁾を用いた提案手法は、

石井らの手法⁽³⁾に比較し、複数等質テストの構成数が少なかったため、出題パターンが限られたものになっている。また、アイテムバンクの項目も十分に活用しているとは言えない。従来の適応型テストは、出題パターンが最も少なく、また、アイテムバンク中で使用された項目が最も少なく、多くの項目を活用しきれていない。この実験の結果により、石井らの手法⁽³⁾を用いた提案手法はアイテムバンクを十分に有効活用しているといえる。

表1 複数等質テストの情報量の上限と下限

情報量関数 (下限/上限)				
$\theta = -2$	$\theta = -1$	$\theta = 0$	$\theta = 1.0$	$\theta = -2$
2/2.4	3.2/3.6	3.2/3.6	3.2/3.6	2/2.4

4. おわりに

本研究では、アイテムバンクを有効活用するため、複数等質適応型テストの出題手法を提案し、シミュレーション実験によって提案手法の有効性を示した。

謝辞

本研究は科研費 (15K21007) の助成を受けたものである。

参考文献

- (1) 植野真臣, 永岡慶三: “e テスティング”, 培風館, 東京 (2009)
- (2) Howard Wainer. Rescuing computerized testing by breaking zipf's law. Journal of Educational and Behavioral Statistics, Vol. 25, pp. 203-224, (2000)
- (3) Ishii, T., Songmuang, P., & Ueno, M. Maximum Clique Algorithm and Its Approximation for Uniform Test Form Assembly. Learning Technologies, IEEE Transactions on, 7(1), 83-95, (2014)
- (4) Van der Linden, W. J. Linear Models for Optimal Test Design (Statistics for Social and Behavioral Sciences), Springer, (2005)

表2 シミュレーション評価の結果

複数等質テストの構成手法		提案手法				従来の適応型テスト
		重複項目数：5		重複項目数：0		
能力	構成された複数等質テスト数	石井らの手法 ⁽³⁾	Linden の手法 ⁽⁴⁾	石井らの手法 ⁽³⁾	Linden の手法 ⁽⁴⁾	
-2	出題パターン数	1000	851	855	885	19
	テストに出現した項目数	1973	778	800	1325	18
-1	出題パターン数	1000	918	912	931	24
	テストに出現した項目数	1966	749	800	1313	23
0	出題パターン数	1000	904	903	932	29
	テストに出現した項目数	1594	548	670	1037	24
1	出題パターン数	1000	903	934	919	24
	テストに出現した項目数	1958	753	800	1286	23
2	出題パターン数	1000	846	867	886	19
	テストに出現した項目数	1973	778	800	1325	18