

平成 30 年 6 月 11 日現在

機関番号：12604

研究種目：若手研究(B)

研究期間：2015～2017

課題番号：15K21007

研究課題名(和文)eテストにおける複数等質適応型テストの自動生成システムの開発

研究課題名(英文)Uniform Adaptive Testing in e-Testing

研究代表者

宮澤 芳光(MIYAZAWA, Yoshimitsu)

東京学芸大学・次世代教育研究推進機構・助教

研究者番号：70726166

交付決定額(研究期間全体):(直接経費) 3,200,000円

研究成果の概要(和文): 適応型テストとは、測定精度を減少させずに受検時間や項目数を軽減させるため、受検者の能力に応じて測定精度が最も高い項目を出題するコンピュータ・テストの出題形式である。しかし、従来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可能性がある。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同一の測定精度を保ちつつ適応的に出題できる等質適応型テストを提案する。具体的には、1) 異なる項目で構成されるが測定精度が等質になるような等質テストを多数構成する。2) 受検者ごとに異なる等質テストを一つ割り当て、そのテスト内の項目集合をアイテムバンクとみなして適応型テストを実施する。

研究成果の概要(英文): Adaptive testing is a question format of computer testing that estimates an examinee's ability sequentially and which selects question items with the highest estimation accuracy according to the examinee's ability. However, in conventional adaptive tests, it is highly likely that the exact same group of items will be selected for examinees who have equivalent ability. We propose a uniform adaptive test that adaptively selects different items for examinees even if those with equivalent capabilities, maintaining the same evaluation accuracy. Specifically, we follow the procedure outlined below. 1) We compose an item cluster for a multiple equivalent test based on the amount of test information. 2) Regarding an item cluster for a multiple equivalent test as an item bank, we propose a uniform adaptive test that estimates the value of an examinee's ability sequentially and which selects items with the highest estimation accuracy.

研究分野：教育工学

キーワード：適応型テスト 複数等質テスト構成 eテストング 項目反応理論

### 1. 研究開始当初の背景

近年、e テスティングと呼ばれる、Web 上でテストを受験する CBT (Computer based testing) の実用化が注目されている。

e テスティングには、任意のタイミングで何度でも同一の測定精度のテストを受検できる利点がある。そのため、テストの結果が受検者に大きな影響を及ぼすハイ・ステークステストを含む様々なテスト場面において、その導入が進みつつある。一方、CBT の出題方式の一つとして適応型テストが知られている。適応型テストでは、受検者の解答のたびに能力を推定し、その能力推定値に対して情報量が高い項目を出題する。このように受検者の能力に応じて項目を逐次的に出題することで、受検者の測定精度を減少させることなく、出題項目数や受検時間を軽減できる利点がある。また、テスト終了基準を適切に設定することで、全ての受検者を同程度の精度で測定することも可能である。しかし、従来の適応型テストには、以下の2つの問題がある。

- (1) 同一の受検者が複数回受験した場合には、同一の項目群が出題される傾向があり、実際に適応型テストを導入している SPI や GTEC の重要な問題になっている。
- (2) 能力が同等な受検者には同一の項目群が出題される可能性が高いため、アイテムバンクの全ての項目を広く一様に活用することができない。特定の項目群の過度な露出は、受検者への項目内容の暴露につながり、テストの信頼性の低下要因となりうる。

適応型テストを運用するためには、難易度や識別力などの特性が既知のテスト項目集合 (アイテムバンクと呼ばれる) を事前に構築する必要があり、特にハイ・ステークステストにおいては、項目の作成に膨大な経済的・時間的コストを要するため、テスト実施においてはアイテムバンク内の全ての項目をできる限り活用することが望ましい。本研究では、これらの問題を解決する新しい適応型テストの枠組みを提案する。

### 2. 研究の目的

本研究では、能力が同等な受検者であっても、同一の測定精度を保ちつつ、異なる項目を適応的に出題できる等質適応型テストを提案する。具体的には、1) テストの長さ、2) テスト間の測定精度が等質であるにも関わらず、3) 異なる項目を出題できる適応型テストを提案する。

### 3. 研究の方法

提案手法の具体的なアプローチは以下の通りである。

- (1) 最大数の複数等質テストを構成できる Ishii et al. (2014) の手法を用いて複数等質テストを構成する。
- (2) 受検者ごとに異なる等質テストを割り当て、そのテスト内の項目集合をアイテム

バンクとみなして適応型テストを実施する。

提案手法では、受検者ごとに異なる項目集合をアイテムバンクとして用いるため、能力が同等な受検者であっても異なる項目群を出題することができる。また、これにより出題される項目出題の多様性が向上するため、アイテムバンク内の項目を満遍なく利用でき、露出数の偏りも軽減されると期待できる。提案手法では、受検者ごとにできる限り異なる等質テストを割り当てる必要があるため、2017年時点で現存する手法で最大の複数等質テストを構成できる Ishii et al. (2014) の手法を用いることが主な提案である。

Ishii et al. (2014) の手法では、テスト構成問題を最大クリーク問題として扱う。具体的には、次のグラフを考え、そこから最大のクリーク (その集合に含まれる任意の頂点がすべて結合されている) 構造を抽出することで複数等質テストを構成する。

**頂点:** 与えられたアイテムバンクから構成可能な、重複条件以外の全てのテスト構成条件を満たす、可能テストを頂点とする。

**エッジ:** 二つの可能テストが重複条件を満たしている場合 (重複条件により指示される最大重複項目数より少ない重複項目しか持っていないなら) その二つの頂点 (テスト) 間にエッジを張る。

このように作成されたグラフのクリークは所望の等質条件を満たした等質テストの集合と解釈できる。

そのため、このグラフの最大クリークを抽出することで、最大数の複数等質テストを生成できる。

提案手法では、Ishii et al. (2014) の手法により生成された複数等質テストの中から、各受検者に異なるテストを一つ割り当て、それに含まれる項目集合をアイテムバンクとみなして適応型テストを行う。このとき、受検者へのテストの割り当てはランダムに行い、一度受検者に割り当てたテストは候補から削除し、次の受検者には残りの候補からランダムに割り当てを行う。候補が空になったら、再度すべてのテストを候補とする。

提案手法では、受検者ごとに異なる項目集合から項目を選択するため、能力が同等な受検者であっても異なる項目が出題できると期待できる。さらに、出題される項目の多様性が向上するため、アイテムバンク内の項目を満遍なく利用することができ、露出数の偏りも軽減されると予想できる。

### 4. 研究成果

本研究では、提案手法の有効性を確認するために、シミュレーション実験による評価を行った。ここでは、通常の適応型テスト (CAT)、制約付き適応型テスト (CCAT と呼ぶ)、アイテムバンク分割法 (IPP と呼ぶ) との比較を行う。シミュレーション実験の手順は以下の通りである。

- (1) 1000 項目で構成されるアイテムバンクを生成する。これらの項目数は、本研究の実データ実験で使用する SPI のアイテムバンクの項目数を基準に決定した。具体的には、SPI の項目数が 978 であったことから項目数 1000 とした。このとき、各項目のパラメータ真値は  $a_i \sim U(0,1)$ ,  $b_i \sim N(0,1)$  からランダムに設定した。
- (2) 受検者の真の能力値を  $\theta \sim N(0,1)$  からサンプリングした。
- (3) 受検者の能力推定値を  $\hat{\theta} = 0$  に初期化した。
- (4) 各手法を用いてアイテムバンクから項目を選択し、その項目への反応データを、能力真値と項目パラメータを所与として発生させた。
- (5) 回答履歴データから能力推定値  $\hat{\theta}$  を EAP 法により推定した。
- (6) 手順 4 と 5 を、推定値の更新幅が  $\varepsilon$  以下になるまで繰り返した。  $\varepsilon$  には、現実の適応型テストで一般に利用される 0.01 を用いた。
- (7) 手順 2 から 6 を 1000 回繰り返し、得られた出題パターンと回答履歴を用いて、次の指標に関する統計量を求めた。
  - a) 適応型テストの長さ、
  - b) 各項目の露出数、
  - c) 異なる受検者に同一でない項目を出題した割合（以降では異なる項目の割合と呼ぶ）、
  - d) 測定精度の等質性

測定精度の等質性については、能力推定値の漸近的な標準誤差を受検者ごとに算出し、それらの値に関する標準偏差の逆数として評価する。なお、標準誤差の平均については、終了条件が全て同じ精度に設定してあるため、各手法の推定誤差はほぼ等しくなっている。ここでは、さらに、実データを用いて提案手法の有効性を評価する。実際に e テスティングを実施しているテスト事業者のアイテムバンクを用いてシミュレーション実験と同様の手順で実験を行った。アイテムバンクの項目数は 978 であった。シミュレーションから生成したアイテムバンクを用いた分析結果を表 1、実際に e テスティングを実施しているテスト事業者のアイテムバンクを用いた分析結果を表 2 に示す。この結果から、以下の特徴が示された。

- (1) テストの長さについては提案手法が最も短かったことが確認できる。 $\theta$  の初期値が受検者の真の能力値から遠い場合、初期値に対して情報量が高い項目を出題するより、情報量が低い項目を出題した方が能力推定の収束が早くなることが知られている。提案手法は、項目数を等質条件で制限するため、ある推定値だけに極端に高い情報量を示す項目を出題しない性質を持つ。この性質により、テストの長さが短縮され、結果として項目の露出数を減少できる。

- (2) 測定精度の等質性は、情報量が高い一部の項目群を繰り返し出題している CAT が最も高かった。一方、CCAT や IPP は、露出数を制御し、受検者ごとに異なる項目を出題しているが、出題された項目に測定精度の偏りがあるため、測定精度の等質性が低い結果となった。提案手法は、測定精度が等質な項目集合から項目を選択しているため、測定精度の等質性は CAT に次いで高く、同一の測定精度を保てたことがわかる。
- (3) 受検者に出題したテスト間で同一ではない項目の割合を表す「異なる項目の割合」を比較すると、提案手法は、異なる項目の割合が最も高くなっている。具体的には、9 割以上の異なる項目を出題していることがわかる。
- (4) 露出数の平均値は、提案手法と IPP が最も低いことが確認できる。提案手法と IPP は、アイテムバンク内の項目を広く一様に活用し、露出数の偏りを減らしていることがわかる。

以上の実験結果から、提案手法の性能について次の点が明らかとなった。提案手法では、受検者に出題される項目のパターンを大幅に増加させることができ、同時にアイテムバンク内の項目をより広く一様に利用できる。加えて、提案手法は、受検者に異なる項目を出題しているにも関わらず、テストの長さや測定精度を等質にすることができた。

表 1 シミュレーションから生成したアイテムバンクを用いた結果

手法	テストの長さ	測定精度の等質性	異なる項目の割合	露出数
CAT	70.4	45.5	0.568	70.4
CCAT	79.5	0.10	0.922	79.5
IPP	56.6	27.8	0.765	56.6
提案手法	35.1	39.61	0.951	35.1

表 2 テスト事業者のアイテムバンクを用いた結果

手法	テストの長さ	測定精度の等質性	異なる項目の割合	露出数
CAT	65.5	62.5	0.512	67.0
CCAT	87.9	3.1	0.911	89.9
IPP	61.7	19.2	0.852	39.7
提案手法	40.7	23.3	0.950	41.6

引用文献

- (1) T. Ishii, P. Songmuang, and M. Ueno (2014) Maximum clique algorithm and its approximation for uniform test form assembly, IEEE Transactions on Learning Technologies, vol.7, no.1, pp.83-95

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 1 件)

- ① 宮澤芳光, 宇都雅輝, 石井隆稔, 植野真臣 (2018) 測定精度の偏り軽減のための等質適応型テストの提案, 電子情報通信学会論文誌 D, vol. J101-D, no. 6, pp. 909-920. 査読有

[学会発表] (計 4 件)

- ① 宮澤芳光, 石井隆稔, 宇都雅輝, 植野真臣 (2017) 整数計画問題を用いた木構造型等質適応型テストの提案, 日本教育工学会第 33 回全国大会.
- ② 宮澤芳光, 石井隆稔, 宇都雅輝, 植野真臣 (2017), 複数等質テストを用いた適応型テスト, 日本テスト学会第 15 回大会発表論文抄録集.
- ③ 宮澤芳光, 石井隆稔, 植野真臣 (2016) 複数等質適応型テストの提案, 日本教育工学会第 32 回全国大会
- ④ 宮澤芳光, 石井隆稔, 植野真臣 (2016) 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, 教育システム情報学会

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

名称 :  
発明者 :  
権利者 :  
種類 :  
番号 :  
出願年月日 :  
国内外の別 :

○取得状況 (計 0 件)

名称 :  
発明者 :  
権利者 :  
種類 :  
番号 :  
取得年月日 :

国内外の別 :

[その他]  
ホームページ等

6. 研究組織

(1) 研究代表者

宮澤芳光 (MIYAZAWA, Yoshimitsu)  
東京学芸大学・次世代教育研究推進機構・  
助教  
研究者番号 : 70726166

(2) 研究分担者

( )

研究者番号 :

(3) 連携研究者

( )

研究者番号 :

(4) 研究協力者

( )