

令和 6 年 6 月 17 日現在

機関番号：82616

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K12170

研究課題名（和文）測定精度の向上のための等質適応型テストの提案

研究課題名（英文）Uniform Adaptive Testing to Improve Measurement Accuracy

研究代表者

宮澤 芳光（Miyazawa, Yoshimitsu）

独立行政法人大学入試センター・研究開発部・准教授

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）：本研究では、露出数の減少と測定精度の増加のトレードオフを制御するため、新しい適応型テストの枠組みを開発した。具体的には、最新の等質テスト構成技術を用いた等質適応型テストを提案した。本手法では、事前にアイテムバンクを分割して情報量が等質な項目集合を複数構成し、テストの前半に項目集合から項目選択し、テストの後半にアイテムバンク全ての項目から項目選択する。本手法の利点として以下の点を確認した。（1）測定誤差を従来の適応型テストと同程度にすることができた（2）項目をできる限り一様に出題でき、項目の暴露数を減少させることができた（3）アイテムバンクの特性やテストの終了条件に関わらず上記の特性を持つ。

研究成果の学術的意義や社会的意義

適応型テストは、受検者の解答から逐次的に能力値を推定し、その能力値に応じて能力推定に最適な項目を出題するテスト形式である。従来の試験と比較し、受検者の能力値を少ない項目で高精度に推定できる。しかし、従来の適応型テストでは、一部の項目群が頻繁に出題される傾向があり、それらの項目群について受検対策されるためにテストの信頼性が低下することが指摘されている。一方で、特定の項目群の頻繁な露出を抑えた場合、測定精度が低下してしまう。つまり、露出数の減少と測定精度の増加にはトレードオフの問題がある。本研究では、露出数と測定精度のトレードオフを制御するため、新しい適応型テストの枠組みを提案している。

研究成果の概要（英文）：In this study, we developed a new adaptive testing framework to control the trade-off between the reduction in the number of exposures and the increase in measurement accuracy. Specifically, we proposed a uniform adaptive test using the latest uniform test assembly techniques. In this method, we first divide the item bank into multiple sets of items with equal information and then select items from these sets in the first half of the test. In the second half of the test, selections are made from all items in the item bank. The advantages of this method were confirmed as follows: (1) it achieved measurement error comparable to conventional adaptive tests, (2) it allowed for the most uniform item exposure possible, thereby reducing the number of item exposures, and (3) it maintained these characteristics regardless of the properties of the item bank or the test termination conditions.

研究分野：教育工学

キーワード：適応型テスト eテスト e テスティング CBT 項目反応理論

1. 研究開始当初の背景

近年、テストの結果が受検者に大きな影響を与えるハイ・ステークスなテストで e テスティングが実用化されつつある。e テスティングとは、Web 上でテストを受ける CBT(Computer based testing) であり、受検者が何度でも等質な測定誤差で異なる項目から構成されたテストを受検できる。e テスティングには、適応型テスト (CAT: Computerized Adaptive Testing) と呼ばれるテスト出題方式が知られている。適応型テストは、受検者の能力値を逐次的に推定し、その能力値に応じて情報量が最大の項目を出題することで、測定誤差を増加させずにテストの長さや受験時間を短縮できる。しかし、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テスト業者の重要な問題になっている。また、特定の項目群が多く受検者に対して提示されてしまうため、別の受検者に共有されてしまうことが指摘されている。これらの問題を解決するため、露出制御を用いた適応型テストが多数提案されている。代表的な手法として、van der Linden(2010)らは、露出数の制約を用いて項目集合(シャドーテストと呼ばれる) を逐次的に構成し、その項目集合から項目選択する手法を提案している。この手法では、項目選択のたびにシャドーテストを構成し、シャドーテスト中の項目から情報量が最大の項目を選択する。一方、別のアプローチとして、確率的に項目選択を制御する手法が提案されている。Sympson-Hetter 手法(Hetter & Sympson 1997)では、事前にシミュレーション実験を用いて項目の出題確率とその出題確率を制御するパラメータを算出し、そのパラメータを用いて確率的に項目の出題を制御する手法が提案されている。これらの手法では、特定の項目群の過度な露出を防ぐことはできる。しかし、これらの手法では、露出数の一様性は担保されるが、情報量を制約しているために測定誤差が増加してしまう問題点がある。すなわち、露出数の減少と測定誤差の増加には、トレードオフの関係がある。

本研究では、この問題を解決するために、露出数の減少と測定誤差の増加のトレードオフを制御する等質適応型テストを提案する。

2. 研究の目的

適応型テストでは、露出数の減少と測定誤差の増加にはトレードオフの関係がある。本研究では、暴露数の減少と測定誤差の増加のトレードオフを制御するために、等質適応型テストを提案する。等質適応型テストでは、最新の等質テスト構成技術を用いて情報量が等質な項目集合にアイテムバンクを分割し、受検者ごとに異なる等質テストを割り当ててアイテムバンクとみなして適応型テストを実施する。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、能力値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、受検者間の測定誤差を等質にしつつ、暴露数を減少できる。等質テストとは、複数のテストが異なる項目で構成されているにも関わらず、等質な情報量を持つ項目集合である。等質テスト構成技術は、既に e テスティング技術の基幹技術として多数研究されている。当初は、その計算量の大きさから少数の等質テストしか生成できなかった。しかし、近年の研究では、大規模な数の等質テストを生成できる技術が開発され、医療系共用試験などの実際のテスト運営でも実用化されている。等質適応型テストでは、アイテムバンクを多数の項目集合に分割し、受検者ごとに異なる項目集合を割り当て、その項目集合から情報量が最大の項目を選択している。これにより、暴露数と測定誤差のトレードオフを制御でき、さらに、等質適応型テストが過学習を避け、テストの長さを短縮できた。しかし、一般的にテストの長さが短縮すると能力推定値の測定誤差が大きくなることが多い。本研

究での分析の結果、アイテムバンクを分割することで項目候補が少なくなり、枯渇して見かけ上能力推定値が収束していた。このため、能力真値と能力推定値が大きく乖離している。ここでは、等質適応型テストの問題を解決するため、2段階等質適応型テストを提案する。

3. 研究の方法

2段階等質適応型テストでは、まず、複数等質テスト構成手法を用いて、情報量や回答所要時間等の統計的性質は等質であるが、異なる項目から構成された等質テストを複数生成する。複数等質テストを生成できる手法としては、石井らの手法が知られている。石井ら(2018)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。さらに、石井ら(2018)の手法では、整数計画問題を用いて最大クリークを逐次探索し、効率的に等質テストを構成できる手法を提案している。

石井ら(2018)の手法では、構成中の等質テスト群を C 、構成済みの等質テスト数を $|C|$ とし、以下の整数計画問題を用いて等質テストを構成する。

$$\text{Maximize } \sum_{i=1}^N \lambda_i x_i$$

Subject to

$$\sum_{i=1}^N y_{ki} x_i \leq 0 \text{ (項目の重複上限数)}; (k = 1, \dots, |C|)$$

$$\sum_{i=1}^N x_i = n \text{ (テストの長さ)}$$

$$\sum_{i=1}^N I_i(\theta_l) x_i = I(\theta_l)$$

$$LB(\theta_l) \leq I(\theta_l) \leq UB(\theta_l)$$

$$(l = 1, \dots, L)$$

$$x_i = \begin{cases} 1: \text{項目 } i \text{ が等質テストに含まれるとき、} \\ 0: \text{上記以外} \end{cases}$$

$$y_{ki} = \begin{cases} 1: i \text{ 番目の項目が } C \text{ 中の } k \text{ 番目の等質テストに含まれるとき、} \\ 0: \text{上記以外} \end{cases}$$

ここで、 λ_i は、互いに独立な $[0, 1)$ の連続一様分布からの乱数であり、本問題が解かれるたびにリサンプリングされる。構成済みの等質テストを C とし、 0 は、重複項目数の上限である $LB(\theta_l)$ は、情報量の下限であり、 $UB(\theta_l)$ が上限である。 $LB(\theta_l)$ と $UB(\theta_l)$ は、アイテムバンクに含まれる項目の特性に応じて適切に設定する必要がある。本研究では、項目の情報量の平均を m_{ib} とし、標準偏差を sd_{ib} 、等質テストのテストの長さを n としたとき、情報量の上限を $(m_{ib} + sd_{ib})n$

とし、下限を m_{ibn} とした。

2段階等質適応型テストでは、事前に石井他(2017)の手法を用いてアイテムバンクを分割して情報量が等質な項目集合を複数構成する。この項目集合を用いた2段階等質適応型テストのアルゴリズムについて詳述する。第1段階では以下のアルゴリズムに従って項目選択する。

1. 項目集合をランダムに割り当てる。
2. 能力推定値を $\hat{\theta} = 0$ に初期化する。
3. 能力推定値 $\hat{\theta}$ を所与として項目集合から情報量が最大となる項目を選択して出題する。
4. 項目への反応データとそれまでの解答履歴から能力推定値 $\hat{\theta}$ を求める。
5. 手順(3)と(4)を $\hat{\theta}$ の更新幅が閾値以下になるまで繰り返す。

次に、第2段階では以下のように出題方略が変更される。

1. 能力推定値 $\hat{\theta}$ を所与としてアイテムバンク全ての項目集合から情報量が最も高い項目を選択する。ハイステークスな試験では、暴露数の上限値が制約として決まっている場合がある。必要に応じて、このステップで暴露数の上限値を制約として組み込む。
2. 項目への反応データとそれまでの解答履歴から能力推定値 $\hat{\theta}$ を求める。
3. 手順(1)と(2)をテストの終了条件まで繰り返す。

従来の適応型テストでは、能力推定値の更新幅が閾値以下になるときをテスト終了条件として設定していた。しかし、本手法では、テスト終了条件を能力推定値の更新幅が閾値以下とした場合、項目集合に情報量の高い項目がなくなり、能力推定値が能力真値に収束する前にテストが終了することがある。そこで本研究では、テストの長さをテスト終了条件とした。

4. 研究成果

本研究では、従来の適応型テストの問題を解決するため、2段階等質適応型テストを提案した。2段階等質適応型テストでは、事前にアイテムバンクを分割して情報量が等質な項目集合を複数構成し、テストの前半に項目集合から項目選択し、テストの後半にアイテムバンク全ての項目から項目選択する。本研究では、特徴を事前に設定したアイテムバンクと実データのアイテムバンクを用いた実験により、2段階等質適応型テストの利点として以下の点が確認できた。

- 2段階等質適応型テストは、測定誤差をCATと同程度にすることができた。
- 2段階等質適応型テストは、項目をできる限り一様に出題でき、項目の暴露数を減少させることができた。
- 2段階等質適応型テストは、アイテムバンクの特性やテストの終了条件に関わらず上記の特性を持つ。

以上の研究の成果は、日本テスト学会で大会発表賞を受賞し、国際会議AIEDに論文が採択され、電子情報通信学会の論文誌に採択されている。

引用文献

- van der Linden, W. J., & Glas, C. A. (2010). Elements of adaptive testing, Springer.
- Hetter, R. D., & Sympson, J. B. (1997). Item exposure control in CAT-ASVAB. In W. A. Sands, B. K. Waters, & J. R. McBride (Eds.), Computerized adaptive testing: From inquiry to operation (pp. 141-144). American Psychological Association.
- 石井隆稔・赤倉貴子・植野真臣 (2018). 複数等質テスト構成における整数計画問題を用いた最大クリーク探索の近似法, 電子情報通信学会論文誌.D, vol. J101, pp. 725-728.

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 宮澤 芳光, 植野 真臣	4. 巻 J106-D
2. 論文標題 高精度能力推定を保証する2段階等質適応型テスト	5. 発行年 2023年
3. 雑誌名 電子情報通信学会論文誌D 情報・システム	6. 最初と最後の頁 34 ~ 46
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2021JDP7052	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Maomi Ueno, Yoshimitsu Miyazawa	4. 巻 13355
2. 論文標題 Two-Stage Uniform Adaptive Testing to Balance Measurement Accuracy and Item Exposure	5. 発行年 2022年
3. 雑誌名 The conference proceedings AIED 2022	6. 最初と最後の頁 626 ~ 632
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-031-36272-9_60	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 宮澤 芳光, 淵本 孝真, 植野 真臣
2. 発表標題 等質テストの最大クリーク並列探索を用いた2段階等質適応型テスト
3. 学会等名 日本行動計量学会
4. 発表年 2022年

1. 発表者名 宮澤 芳光, 淵本 孝真, 植野 真臣
2. 発表標題 等質テスト構成の並列化技術を用いた2段階等質適応型テスト
3. 学会等名 日本テスト学会
4. 発表年 2022年

1. 発表者名 宮澤芳光, 植野真臣
2. 発表標題 項目暴露を軽減する二段階等質適応型テストの提案
3. 学会等名 日本行動計量学会
4. 発表年 2021年

1. 発表者名 宮澤芳光, 植野真臣
2. 発表標題 最大暴露数を制約に用いた二段階等質適応型テスト
3. 学会等名 日本テスト学会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関