

出題頻度の偏りと測定精度のトレードオフを制御する ZDD を用いた二段階等質適応型テスト

Two-Stage Uniform Adaptive Testing using Zero-suppressed Binary Decision Diagrams to balance Measurement Accuracy and Bias of Item Exposure

淵本 吉真*1, 植野 真臣*1

Kazuma FUCHIMOTO*1, Kazuma FUCHIMOTO*1

*1 電気通信大学大学院 情報理工学研究所

*1 Graduate School of Informatics and Engineering, University of Electro-Communications

Email: fuchimoto@ai.lab.uec.ac.jp

あらまし：適応型テストは、受検者能力を逐次的に推定し、その能力に応じて出題する。しかし、適応型テストは、良問が過度に出題されるため、出題頻度に偏りが生じ、テストの信頼性が低下する。本研究では、ZDD を用いて生成した等質テスト群から異なるテストを受検者に割り当て適応型テストを行う。また、受検者能力の推定誤差が小さくなる後半は、全問題から出題する。提案手法は、測定精度を低下させず、出題頻度の偏りを抑制できた。

キーワード：テスト理論、項目反応理論、Computer Based Testing、適応的学習支援、グラフ理論

1. はじめに

適応型テストでは、受検者の能力を逐次的に推定し、その能力に応じた問題項目を出題する。これにより、適応型テストは、従来の試験よりも少ない問題項目で受検者の能力を高精度に測定できる。しかし、適応型テストは、良問が過度に出題されるため、出題頻度に偏りが生じ、テストの信頼性が低下する。この問題を解決するために、Kishida et al. ⁽¹⁾から着想を得て、本研究では、Zero-suppressed Binary Decision Diagram(ZDD)を用いて生成した等質テスト群から異なるテストを受検者に割り当て適応型テストを行う。また、受検者能力の推定誤差が小さくなる後半は、全問題から出題する。提案手法は、測定精度を低下させず、出題頻度の偏りを抑制できた。

2. 項目反応理論

適応型テストでは、一般的に項目反応理論を用いて受検者能力を測定する。現在、最もよく使われる3母数ロジスティックモデル(3PLM)では、能力値 $\theta_j \in (-\infty, \infty)$ を持つ受験者 $j(=1, 2, \dots, J)$ が項目 $i(=1, 2, \dots, n)$ に正答する確率 $p_i(\theta_j)$ を以下に定義する。

$$p_i(\theta_j) = c_i + \frac{1 - c_i}{1 + \exp(-1.7a_i(\theta_j - b_i))}$$

ただし、 $a_i \in [0, \infty]$, $b_i \in [0, \infty]$, $c_i \in [0, 1]$ はそれぞれ識別力パラメータ、困難度パラメータ、当て推量パラメータと呼ばれる項目パラメータであり、事前に推定された値が問題バンクに格納されている。

また、多値型項目反応モデルとして、一般化部分採点モデル(GPCM)が知られている。GPCMでは、受験者 j が項目 i に対してカテゴリ $k(=1, 2, \dots, K)$ と反応する確率 $p_i(k|\theta_j)$ を以下のように定義する。

$$p_i(k|\theta_j) = c_i + \frac{\exp(\sum_{s=1}^k a_i(\theta_j - b_{i,s}))}{\sum_{k=0}^{K-1} \exp(\sum_{s=1}^k a_i(\theta_j - b_{i,s}))}$$

3PLM および GPCM における受検者能力の推定値の標準誤差は、それぞれ以下のフィッシャー情報量の逆数に漸近的に一致する。

$$I_i(\theta) = \frac{\left[\frac{\partial}{\partial \theta} p_i(\theta) \right]^2}{p_i(\theta)(1 - p_i(\theta))} \quad (3PLM),$$

$$I_i(\theta) = \sum_{k=0}^K \frac{\left[\frac{\partial}{\partial \theta} p_i(k|\theta) \right]^2}{p_i(k|\theta)} \quad (GPCM).$$

フィッシャー情報量 $I_i(\theta)$ の高い項目は、その能力値付近の受検者をよく識別できる。したがって、従来の適応型テストでは、その時点での能力推定値を所与として、フィッシャー情報量最大の項目を出題する。このように能力推定値に対して情報量最大の項目を逐次出題することで、少ない項目数で高精度な能力推定を実現する。

3. 提案手法

適応型テストの貪欲的な出題は、出題頻度に偏りを生じさせ、テストや問題の信頼性が低下する。この問題を解決するために、Kishida et al. は、最大クリークと整数計画法を用いた難易度制約付きの二段階等質適応型テスト手法を提案している⁽¹⁾。しかし、この手法では3つの課題がある。(1) 出題頻度の偏りと測定精度のトレードオフが十分に解消されていない。(2) GPCMについて、難易度制約を課していない。(3) 複数回受検時に、同じ問題項目が提供される可能性がある。

これらの問題を解決するために、本研究では、Kishida et al. から着想を得たZDDを用いた難易度制約付きの二段階等質適応型テストを提案する。初めに、提案手法では、ZDDを用いた自動平行テスト構成⁽²⁾を用いて、受検者能力の測定精度が等質な問題項目の組合せを列挙する。次に、第一段階目では、

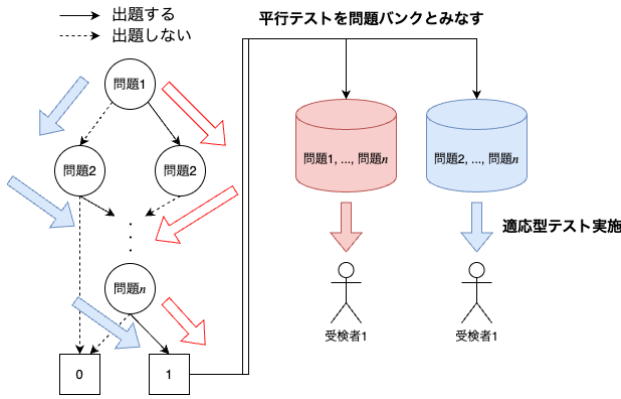


図 1 ZDD を用いた提案手法の第一段階概要

図 1 のように、ZDD を用いた平行テストからランダムサンプリングを行い、適応型テストを行う。この際、複数回受検者に対しては、前回受検時の項目を ZDD の集合演算を用いて除外できるため、(3)の問題を解消できる。その後、受検者能力の更新幅が小さくなると、第二段階目に切り替わる。二段階目では、能力推定値近傍の難易度パラメタ区間に属する項目から適応的項目出題を行う。3PLM の項目については、Kishida et al.と同様に以下の条件を満たし、フィッシャー情報量が最大となる項目を問題バンク全体から選択・出題する。

$$\hat{\theta} - \delta SD(\hat{\theta}) < b < \hat{\theta} + \delta SD(\hat{\theta}).$$

ここで、 $SD(\hat{\theta})$ は能力推定値の事後標準偏差、 δ はチューニングパラメタである。さらに、本研究では、(2)の問題を解消するために、GPCM のための難易度パラメタ区間を以下に定義する。

$$\hat{\theta} - \delta SD(\hat{\theta}) < b_{i,s} < \hat{\theta} + \delta SD(\hat{\theta}).$$

ただし、 $\exists s \in \{1, 2, \dots, K\}$ が存在するとき、出題が可能である。

4. 比較実験

ここでは、提案手法を従来手法 (CAT, Hybrid⁽³⁾, HMCAPIP⁽¹⁾) と比較した。実験設定については、Kishida et al. ⁽¹⁾と同様に 30 問の適応型テストをシミュレーションした。また、実験には、以下に示すシミュレーション問題バンクと実データ問題バンク (SPI⁽⁴⁾, Science⁽³⁾) を用いた。結果を表 1 に示す。ここで、RMSE (受検者能力の測定精度)、SD.IER (出題頻度の標準偏差)、Max.IER (最大出題率) は以下の式で計算される。

$$RMSE = \sqrt{\sum_{j=1}^J (\theta_j - \hat{\theta}_j)^2},$$

$$SD.IER = \sqrt{\frac{1}{n} \sum_{i=1}^n (IER_{i,j} - IER_{\mu})^2},$$

$$IER_{i,j} = \frac{1}{J} \sum_{j=1}^J IE_{i,j},$$

$$IE_{i,j} = \begin{cases} 1 & \text{受検者 } j \text{ に項目 } i \text{ が出題された場合} \\ 0 & \text{それ以外} \end{cases},$$

$$Max.IER = \max(\{IER_{1,j}, IER_{2,j}, \dots, IER_{n,j}\}).$$

ただし、 IER_{μ} は、全問題目における $IER_{i,j}$ の平均値である。また、未出題については、問題バンクから出題されなかった問題項目の総数である。

表 1 より、提案手法は、ほぼ全ての条件において、従来手法よりも受検者能力の測定精度を低下させることなく、出題頻度の偏りや最大出題率を抑えることができた。これにより、(3)の問題を従来手法よりも改善できた。また、提案手法は、未出題数も最も少なく、問題バンクの項目を有効活用できた。また、紙面の都合上詳細は割愛するが、以下の 4 点を実験により確認できた。

1. GPCM の難易度制約による出題頻度偏り軽減の効果
2. 項目パラメタと出題頻度の関係性の分析と提案手法の効果
3. 提案手法におけるパラメタチューニング分析
4. 問題数を固定しない場合の提案手法の効果

表 1 提案手法と従来手法の比較結果

問題バンク	手法	RMSE	SD.IER	Max.IER	未出題数
Sim 1	CAT	0.26	0.105	1.000	843
	Hybrid	0.26	0.085	0.495	542
	HMCAPIP	0.26	0.066	0.434	219
	Proposed	0.26	0.061	0.372	60
Sim 2	CAT	0.26	0.116	1.000	874
	Hybrid	0.26	0.095	0.500	689
	HMCAPIP	0.26	0.071	0.557	281
	Proposed	0.26	0.056	0.365	52
SPI	CAT	0.26	0.114	1.000	850
	Hybrid	0.27	0.095	0.500	678
	HMCAPIP	0.26	0.071	0.593	398
	Proposed	0.26	0.056	0.531	281
Science	CAT	0.21	0.122	1.000	893
	Hybrid	0.22	0.086	0.496	689
	HMCAPIP	0.24	0.071	0.529	399
	Proposed	0.24	0.066	0.487	236

5. まとめ

本研究では、Zero-suppressed Binary Decision Diagram(ZDD)を用いた二段階等質適応型テストを提案した。結果として、提案手法は

今後は ZDD の構造上で適応型テストの出題・回答パターンを列挙できる手法への拡張を目指す。

6. 参考文献

参考文献

- (1) Kishida, Wakaba, et al. "Item Difficulty Constrained Uniform Adaptive Testing." International Conference on Artificial Intelligence in Education. Cham: Springer Nature Switzerland, 2023.
- (2) Fuchimoto, Kazuma, Shin-ichi Minato, and Maomi Ueno. "Automated Test Assembly using Zero-suppressed Binary Decision Diagrams." IEEE Access (2023).
- (3) S. Lim and S. Choi, "Item exposure and utilization control methods for optimal test assembly," Behaviormetrika, 2023.
- (4) Recruit, "Synthetic Personality Inventory". <https://www.recruit-ms.co.jp/freshers/spi-001.html>. アクセス日時: 2024年6月1日