

# 自動テスト構成における 項目露出の偏りを軽減する 整数計画法を用いた最大クリーク探索

## Maximum Clique Algorithm using Integer Programming for Reducing Item Exposure Bias in Automated Test Assembly

淵本 亮真<sup>\*1</sup>, 植野 真臣<sup>\*1</sup>

Kazuma FUCHIMOTO<sup>\*1</sup>, Maomi UENO<sup>\*1</sup>

<sup>\*1</sup>電気通信大学大学院 情報理工学研究所

<sup>\*1</sup>Graduate School of Informatics and Engineering, The University of Electro-Communications

あらまし：近年、e テスティングと呼ばれる、Web 上でテストを受検する Computer Based Testing が普及している。e テスティングでは、異なるテストを受検したにも関わらず何度でも同一精度の測定を実現できる。最先端の自動テスト構成手法では項目の出題頻度に偏りが生じる。このような項目は受験対策により受検者の予測測定誤差の劣化が進む。この問題を解決するために、本研究では出題頻度の偏りを軽減するための自動テスト構成手法を提案する。

キーワード：e テスティング，自動テスト構成，並行テスト，テスト理論，項目反応理論

### 1. はじめに

e テスティングとは、異なる問題で構成されるテストを受検したにも関わらず、何度でも同一精度の測定を実現できる。国内では、医療系大学間共用試験などで既に採用されている。e テスティングでは、テストの問題（以降、項目と呼ぶ）が格納されたデータベースから多数の受検者の予測測定誤差を可能な限り小さくかつ等質にテスト群を構成する。近年では、項目反応理論を用いた自動テスト構成手法<sup>(1)~(4)</sup>が数多く提案されている。自動テスト構成の重要な課題の一つは可能な限り多くのテストを生成することである。この課題を解決するために、Ishii ら (2017) は最大クリーク探索と整数計画法を用いた手法を提案した<sup>(3)</sup>。さらに、Fuchimoto ら (2022) は最大クリーク探索と整数計画法の手法が持つ時間・空間計算量の長所を融合し、並列探索可能な二段階並列探索手法を提案した<sup>(4)</sup>。本手法は世界で最も多くのテストを構成できる。しかし、これらの手法<sup>(3)</sup><sup>(4)</sup>はテスト間に項目の重複を許すため、各項目の出題頻度（以降、露出数と呼ぶ）に偏りが生じる。これにより、露出率の高い項目は受験対策などにより受検者の予測測定誤差の劣化が進む問題がある。この問題を解決するために、本研究では項目の出題頻度の偏りを軽減するための自動テスト構成手法を提案する。具体的には、露出数の低い項目ほど優先してテストに含まれる確率が高くなるように整数計画問題の目的関数を定式化し、露出数の偏りを軽減する。

### 2. 項目反応理論

Lord and Novick (1968) は同じ真の得点を測定する2つのテストについて、並行テストと定義した<sup>(5)</sup>。しかし、並行テストは古典的テスト理論における仮定であり、このようなテストの実現は困難である。そのため、Samejima ら (1977) は項目反応理論 (Item

Response Theory: IRT) を用いて、並行テストの概念を拡張した<sup>(5)</sup>。IRT で最もよく使われる2母数ロジスティックモデルでは、能力値 $\theta_j \in (-\infty, \infty)$ を持つ受検者 $j$ が項目 $i$ に正答する確率 $p_i(\theta_j)$ を以下のように定義する。

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

ただし、 $a_i \in [0, \infty]$ ,  $b_i \in [0, \infty]$ は識別力パラメータ、困難度パラメータと呼ばれる項目パラメータである。

また、項目 $i$ において、確率 $p_i(\theta_j)$ のフィッシャー情報量を項目情報量 $I_i(\theta)$ と呼び、以下のように表す。

$$I_i(\theta) = 1.7^2 a_i^2 p_i(\theta)(1 - p_i(\theta))$$

さらに、テストに含まれる項目の項目情報量の総和をテスト情報量 $I(\theta)$ と呼び、以下のように表す。

$$I(\theta) = \sum_{i \in T} I_i(\theta)$$

ここで、 $T$ はテストに含まれる項目の集合である。このテスト情報量の逆数が受検者の能力推定値の漸近分散に収束する。Samejima ら (1977) はこのテスト情報量が等価なテストを弱並行テストとして定義した<sup>(6)</sup>。しかし、能力値 $\theta_j \in (-\infty, \infty)$ について、テスト情報量が等価なテストを生成することは困難なため、多くの先行研究は受検者の能力値を $\theta_k = (\theta_1, \theta_2, \dots, \theta_k)$ と離散的に取り扱う。例えば、Ishii ら (2014) は $\theta_k$ におけるテスト情報量の上限 $UB_{\theta_k}$ ・下限 $LB_{\theta_k}$ 制約の範囲に収まるテストを受検者の予測測定誤差が等質であるとした<sup>(6)</sup>。

### 3. 先行研究

#### 3.1 最大クリーク探索による自動テスト構成

Ishii ら (2014) は自動テスト構成をグラフ上で定義される最大クリーク問題に帰着する手法を提案した<sup>(6)</sup>。本手法では、図1のように、テストをグラフ上の頂点とみなし、等質な場合に辺を引く。このグラフから最大クリーク（任意の二頂点が隣接しているグラフ構造）を抽出することでテストを構成する。

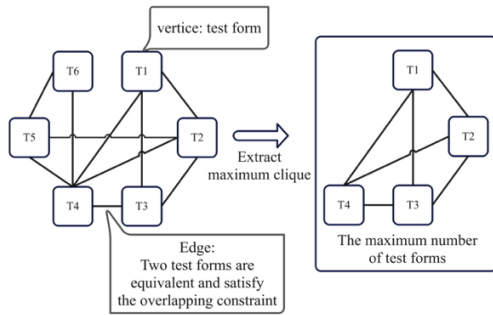


図1 最大クリーク探索による自動テスト構成

### 3.2 整数計画法を用いた最大クリーク探索による自動テスト構成の並列化

Ishii ら (2014) の手法では、空間計算量が $O(|V|^2)$ と大きく ( $V$ は頂点集合)、テスト構成数の改善に限界があった。この問題を解決するために、Ishii ら (2017) は整数計画問題により逐次的にテストを生成することで空間計算量 $O(|V|)$ に減少させた<sup>(3)</sup>。さらに、Fuchimoto らは最大クリーク探索手法の時間計算量が整数計画問題の時間計算量よりも低いことを利用した二段階並列探索アルゴリズムを提案した<sup>(4)</sup>。

### 4. 提案手法

前述の手法<sup>(3)</sup><sup>(4)</sup>は世界で最も多くのテストを構成可能であるが、各項目の露出数に偏りが生じる。特に、最大クリーク探索手法による露出数の偏りが大きく、改善の余地がある。本研究では、この問題を解決するために、最大クリーク探索を行うグラフにおける頂点の生成を整数計画法により逐次的に行う。具体的には、下記の整数計画問題を解くことで逐次的に頂点を生成する。

$$\text{Maximize } \sum_{i=1}^n \left( \lambda_i + \frac{1}{1 + e^{-z_i}} \right) x_i$$

$$z_i = \frac{\sum_{v=1}^{|V|} X_{(i,v)} - \frac{1}{n} \sum_{j=1}^n \sum_{v=1}^{|V|} X_{(j,v)}}{\sqrt{\frac{1}{n} (\sum_{v=1}^{|V|} X_{(i,v)} - \sum_{j=1}^n \sum_{v=1}^{|V|} X_{(j,v)})}}$$

$$\text{Subject to } \sum_{i=1}^n x_i = M, LB_{\theta_k} \leq \sum_{i=1}^n I_i(\theta_k) x_i \leq UB_{\theta_k}$$

ここで、 $x_i$ は項目 $i$ をテストに含む場合に1、それ以外は0をとる決定変数であり、 $\lambda_i$ は互いに独立な $[0,1]$ からの連続一様乱数である。また、 $M$ はテスト項目数、 $LB_{\theta_k}$ 、 $UB_{\theta_k}$ は $\theta_k$ におけるテスト情報量の下限・上限制約、 $V$ は既に生成した頂点の有限集合、 $X_{(i,v)}$ は頂点集合 $V$ における $v$ 番目の要素(テスト)に項目 $i$ 含まれる場合に1、それ以外の場合に0をとる。

### 5. 評価実験

提案手法の有効性を示すために、実データ<sup>(4)</sup>を用いて HMCAPIP 法<sup>(4)</sup>とテスト構成数及び露出数を比較する。結果は表1の通りである。ここで、No. testsはテスト構成数、Item Exposureの各列(Max, Min, SD)はそれぞれ露出数の最大値、最小値、標準偏差を示している。

表1 従来手法との比較実験

OC	HMCAPIP (first step)				Proposal (first step)			
	Item Exposure				Item Exposure			
	No. tests	Max	Min	SD	No. tests	Max	Min	SD
5	45,677	5,126	0	364.8	56,051	2,691	983	87.7

表1より、提案手法は従来手法よりも露出数の標準偏差が小さく、偏りの小さいテスト群を生成できた。また、従来手法よりも最大露出数の最大値が小さく、最小値が大きい。ゆえに、提案手法は過度な項目の出題を防ぎ、従来手法では余り活用されていなかった項目を活用できる。図2は各項目の露出数をプロットしたものである。具体的には、横軸に困難度パラメータ、縦軸にその項目の露出数を示している。図2より、提案手法は従来手法の偏りを抑制し、露出数が均一なことがわかる。

提案手法は、偏りを抑えるだけではなく、過度な項目の出題を防ぎ、従来手法では余り活用されていなかった項目をテストに追加できた。特に、アイテムバンクの項目は作問するためのコストが高く非常に貴重なため、e テスティング運用の敷居を下げるためにも非常に有用性が高いと考える。

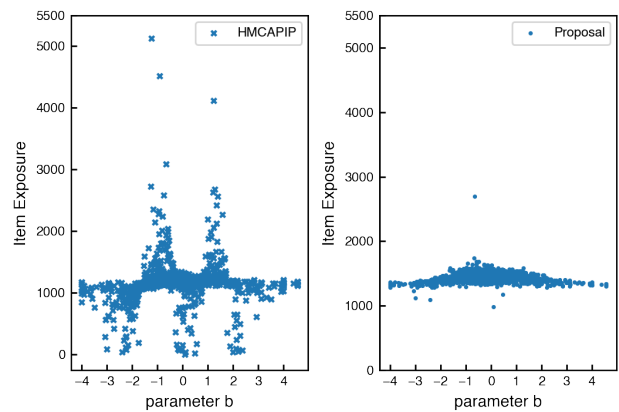


図2 困難度パラメータと露出数の分布

### 参考文献

- (1) W.J. van der Linden, *Liner Models for Optimal Test Design*, Springer, 2005.
- (2) Xiao Luo. Automated test assembly with mixed-integer programming: The effects of modeling approaches and solvers. *Journal of Educational Measurement*, Vol. 57, No. 4, pp. 547-565, 2020.
- (3) Takatoshi Ishii and Maomi Ueno. Algorithm for uniform test assembly using a maximum clique problem and integer programming. In *International Conference on Artificial Intelligence in Education*, pp.102-112. Springer, 2017.
- (4) Kazuma Fuchimoto, Takatoshi Ishii, and Maomi Ueno. Hybrid maximum clique algorithm using parallel integer programming for uniform test assembly. *IEEE Transactions on Learning Technologies*, (2022). (Early Access)
- (5) Samejima, F.: Weakly parallel tests in latent trait theory with some criticisms of classical test theory, *Psychometrika*, Vol. 42, No. 2, pp. 193-198, (1977).
- (6) Ishii, Takatoshi, Pokpong Songmuang, and Maomi Ueno. "Maximum clique algorithm and its approximation for uniform test form assembly." *IEEE Transactions on Learning Technologies* 7.1 (2014): 83-95.