

項目露出を考慮した整数計画法による等質テスト構成

植野 晶^{†a)} 渕本 竜真^{†b)} 植野 真臣^{†c)}

Uniform Test Assembly with Lower Item Exposure Using Integer Programming

Akira UENO^{†a)}, Kazuma FUCHIMOTO^{†b)}, and Maomi UENO^{†c)}

あらまし 本論文では、項目露出を考慮した整数計画法による等質テストの自動構成手法を提案する。等質テストとはテストに含まれる項目は異なるが、受験者得点の予測誤差が等質なテスト群である。等質テストでは同一能力の受験者ならば、どのテストを受験しても同一得点となる保証があり、出題可能な項目のデータベース（アイテムバンク）から自動構成される。項目の露出数（出題回数）が大きい項目は受験者間で共有されやすく、経年劣化につながりその項目の信頼性が失われやすいため、露出数の偏りの軽減が重要な課題の一つである。Ishii and Ueno (2015) は構成した等質テストを全て保存し、その中から最も露出率（＝露出数の最大値/テスト構成数）が小さい等質テストを出力して、露出率を軽減する手法を提案した。本論文では乱数を用いた整数計画法でテスト構成することで、露出数の偏りが改善されることを示し、整数計画法のみでは解決されなかった項目露出のバイアス問題を露出数上位の項目をアイテムバンクから除外しながらテスト構成することで解決する手法を提案する。評価実験では従来手法とテスト構成数及び露出率を比較し、提案手法の有効性を示す。

キーワード e テスティング, 等質テスト自動構成, 最大クリーク問題, 整数計画法, 項目露出

1. ま え が き

e テスティングとは、異なる問題から構成され、同一精度の測定を実現できるコンピュータテストのことである。e テスティングを用いることで、同一能力の受験者が異なるテストを受験しても同一得点となる保証がある。そのために、受験者が同一精度で複数回の受験が可能となる [1]～[3]。e テスティングではテストに含まれる項目は異なるが、出題項目数や得点の予測誤差がテスト間で等しいテスト群、等質テストを用いる。等質テストにより異なる項目から構成されたテストを同一能力の受験者が受験した場合同一得点になることが保証される。これまでに、等質テストを自動構成する手法が数多く提案されている [4]～[9]。

一般に e テスティングでは、アイテムバンクと呼ばれる出題項目を管理するデータベースが利用される。アイテムバンクには出題項目の統計データが格納され

ており、所望のテストの性質を満たす組み合わせを計算機で自動構成する。図 1 はテストの自動構成の概念図である。

テストの自動構成は数理最適化問題として解かれる。例えば、Songmuang and Ueno (2010) は最適化問題の解探索手法の一つである Bees Algorithm を用いてテスト構成を提案した。この手法は我が国の国家試験である情報処理技術者試験で実際に使用されている [8]。

Ishii, Songmuang, and Ueno (2014) はグラフ上で定義される最大クリーク問題に帰着してテスト構成を行う手法を提案した。具体的には与えられたアイテムバンクからテスト構成条件を満たすテストを生成し、グラフの頂点とみなす。二つのテスト（頂点）が等質条件を満たす場合に辺を引いたグラフ構造から最大クリー

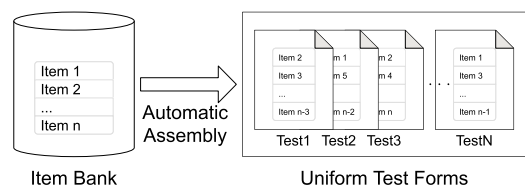


図 1 テストの自動構成
Fig. 1 Automatic assembly for test forms.

[†] 電気通信大学大学院情報システム学研究科, 調布市
Graduate School of Information Systems, The University of Electro-Communications, Chofu-shi, 182-8585 Japan

a) E-mail: ueno.akira@ai.lab.uec.ac.jp

b) E-mail: fuchimoto@ai.lab.uec.ac.jp

c) E-mail: ueno@ai.is.uec.ac.jp

DOI: 10.14923/transinfj.2021JDP7037

ク探索することで等質テストを構成する [10], [11]. この手法は理論的に最大のテスト構成を保証するが, アイテムバンクの項目数に対して構成可能な頂点 (テスト) 数が組み合わせ爆発的に増加するため, 最大クリーク探索が困難である. この問題を解決するために, Ishii, Songmuang, and Ueno (2014) はグラフからランダムに部分グラフを選択し, 最大クリーク探索を繰り返すことによりグラフ全体の最大クリークを近似的に探索する手法 (RndMCP 法) を提案した. 本手法により, 当時の既存研究よりも 10~100 倍以上多くのテストを構成している [12]. 最大クリーク探索はグラフの頂点集合を V とすると, 最先端の最大クリーク探索手法 [13], [14] を用いても $O(|V|^2)$ の空間計算量を必要とするため, 最大で 10 万程度のテスト構成が限界であった.

RndMCP 法の空間計算量を緩和するために, Ishii and Ueno (2017) は第 1 段階目で RndMCP 法を用いてメモリの限界まで大きな最大クリークを探索した後, 第 2 段階目で第 1 段階目で求めたクリークの全頂点と隣接する頂点を整数計画法により逐次的に探索することで空間計算量を $O(|V|)$ に削減させる HybridRBP 法を提案した [15], [16]. これにより 10 万を超える等質テストを構成することが可能になった. しかし, 整数計画法の時間計算量が $O(2^n)$ (n はアイテムバンクの項目数) と大きく, テスト構成数の改善はわずかであった. そのため, 渕本, 植野 (2020) は HybridRBP 法 [15], [16] を並列探索アルゴリズムに改善し, 多くのテスト構成を可能とした [17].

先行研究 [12], [15], [16] では構成されたテスト群は項目の露出数 (出題回数) に偏りが生じる問題がある. 露出数が多い項目は受験者間で共有されやすく, 経年劣化につながり, その項目の信頼性が失われやすい [18]. この露出数の偏りを軽減するために, Ishii and Ueno (2015) は RndMCP 法と整数計画法を用いてテスト構成し, その中から最も露出率 (= 露出数の最大値 / テスト構成数) が小さいテスト構成を選択する手法を提案した [19]. 具体的には, 探索した全ての等質テスト群 (クリーク) を候補として保存しておき, 最後にその候補で最も露出率が小さい等質テストを出力する. これによって, 従来手法よりも露出率を軽減することができた. しかし, Ishii and Ueno (2015) を用いても, 特定の項目が過度に露出することが多い.

従来手法 [19] の第 1 段階目の RndMCP 法はランダム探索による最大クリーク探索で, 探索範囲も限定的

になり頂点集合における露出数に偏りが生じる問題がある. この問題を解決するために, 本論文では整数計画法を用いて頂点生成する. 整数計画法を用いることで, 制約条件を満たす項目集合から一様に頂点を生成することができる. これにより, 第 1 段階目の頂点生成において項目露出の偏りを抑制する.

しかし, この整数計画法は制約を満たす解から一様に解を抽出するだけで, 項目露出のバイアス問題は完全には解決されない. そこでこの問題を解決するために, 露出数上位 s 位までの項目をテスト構成ごとにアイテムバンクから除外し, テスト構成を行う手法を提案する. これにより, 露出数上位の項目を除いたアイテムバンクからテスト構成するので, 最大露出数が抑制される. テスト構成ごとにアイテムバンクから項目を除外していくと, いずれは構成条件を満たすようなテストを構成できなくなるが, その際に除外した項目を全てアイテムバンクに戻す. また, パラメータ s はテスト構成数と露出率に影響を与えるのでデータから決定する.

本論文では, 提案手法の有効性をシミュレーション及び実データを用いて示した. 具体的には, 従来手法と比較して最大露出数を抑制し露出率を改善したことが示された. また, 幾つかの条件において, 提案手法では従来手法よりも密なグラフ構造から最大クリーク探索を行うことができた. これにより, 露出率だけでなく, テスト構成数も改善することができた.

2. 項目反応理論

一般的に等質テストは, 以下の構成条件を満たすテストの集合として定義する. (例えば [5], [11], [12])

- (1) それぞれのテストでの受験者得点の予測誤差が等質である.
- (2) それぞれのテスト間の項目重複数が一定値以下である. (以降, 項目重複数条件と呼ぶ)

受験者得点の予測誤差は従来研究 (例えば [4], [8], [9], [20]~[22]) において項目反応理論 (Item Response Theory: IRT) [23], [24] におけるテスト情報量の逆数で推定されている. IRT とは受験者の項目への正答確率をモデル化したものである. これにより, 異なる項目から構築されるテストを受けた受検者の能力を同一尺度上で評価できる.

IRT の中で最もよく使われる 2-パラメータロジス

ティックモデル (2-Parameter Logistic Model: 2PLM) では能力値 $\theta_j \in (-\infty, \infty)$ をもつ受験者 $j = (1, \dots, m)$ が項目 $i = (1, \dots, n)$ に正答する確率 $p_i(\theta_j)$ を以下のようモデル化する。

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

ここで、 $a_i \in (0, \infty)$, $b_i \in (-\infty, \infty)$ はそれぞれ i 番目の項目の識別力パラメータ、困難度パラメータと呼ばれる項目パラメータである。

IRT では項目 i において式 (1) を用いて計算したフィッシャー情報量を項目情報量 $I_i(\theta)$ と呼び、以下のように定義する。

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

また、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と呼び、以下のように表す。

$$I(\theta) = \sum_{i \in T} I_i(\theta) \quad (3)$$

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験者能力推定値の漸近分散に収束することが知られている [1]。

ただし、テストの自動構成手法 (例えば、[4], [8], [20] ~ [22], [25]) ではテスト情報量における受験者の能力パラメータ θ を $\Theta = \{\theta_1, \dots, \theta_k, \dots, \theta_K\}$ のように幾つかの点でサンプリングし、離散的に扱っている。

3. 等質テストの自動構成アルゴリズム

3.1 等質テストのための最大クリーク問題

Ishii et al. (2014) は等質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させる手法 (ExMCP 法) を提案した [11]。ここでクリークはグラフの任意の 2 頂点が隣接する頂点集合である。

本手法では能力値パラメータ $\Theta = \{\theta_1, \dots, \theta_k, \dots, \theta_K\}$ をサンプリングし、各点ごとにテスト情報量 $I(\theta_k)$ の上下限 ($UB(\theta_k)$, $LB(\theta_k)$) を設定し、全ての上下限制約を満たすテストを受験者得点の予測誤差が等質であるとする。

例えば図 2 は表 1 のテスト情報量の上下限制約を与えたときの概念図である。また、図中の #1~#4 は構成されたテストの情報量関数を描いたものである。図 2 のテスト #1 や #2 はテスト情報量の上下限制約を満たすため等質である。一方で、テスト #3 や #4 は上下限

表 1 テスト情報量の上下限制約の例
Table 1 Example of test information constraint.

$\theta = -2.0$	$\theta = -1.0$	$\theta = 0.0$	$\theta = 1.0$	$\theta = 2.0$
0.0/0.2	0.1/0.3	0.1/0.3	0.1/0.3	0.0/0.2

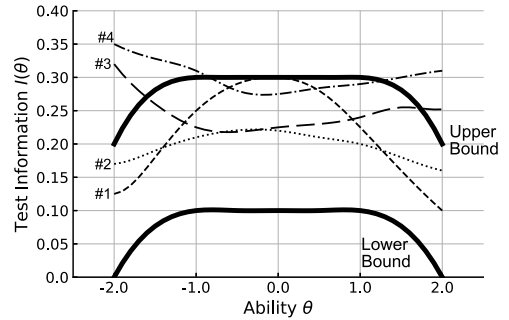


図 2 テスト情報量の上下限の例
Fig. 2 Example of upper and lower bound for test information.

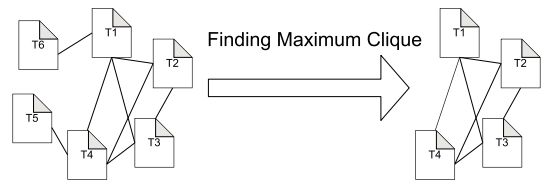


図 3 等質テスト構成のためのクリーク探索
Fig. 3 Clique finding for uniform test assembly.

制約を満たさないため等質でない。

生成されるテスト候補を以下のグラフ構造とみなし、グラフ構造の中からクリークを探索する。

頂点： 項目重複数条件を除いたテスト構成条件を満たす、全てのテストを頂点とする。

辺： 二つの頂点に対応するテストが項目重複数条件を満たす場合、二つの頂点間に辺を引く。

このグラフがもつクリーク中の任意の二頂点は隣接しているため、テスト構成条件を満たす。したがってクリーク中の頂点に対応するテストはそれぞれ等質であり、頂点数が最大のクリークが最大の等質テストとなる (図 3)。

したがって、無向グラフ $G = (V, E)$ の頂点の集合を V 、辺の集合を E としたとき、最大クリーク C (等質テスト群) は次のように定式化できる。

複数等質テスト構成のための最大クリーク問題

variables $C \subseteq V$

maximize $|C|$

subject to

$$\forall v, \forall w \in C, \{v, w\} \in E$$

*ここで $\{v, w\} \in E$ は頂点の組 v, w が
次の条件で引かれた辺を意味する
($|v \cap w| \leq$ 重複項目数の上限値)

この手法は理論的に最大数の等質テストを構成することが可能であるが、最先端の最大クリーク探索手法 [13], [14] を用いても、時間計算量 $O(2^{|V|})$ 、空間計算量 $O(|V|^2)$ を必要とする。等質テストの場合、頂点数 $|V|$ はテスト構成条件を満たすテストの総数となるが、その数はアイテムバンクのテスト項目数 n と比較して組み合わせ爆発的に増加する。ゆえに、現在実施されているような数百〜千以上のアイテムバンクから等質テストの構成を厳密に行うことは困難である。

ExMCP 法の計算コストの問題を緩和するため、Ishii, Songmuang, and Ueno (2014) はグラフ全体の最大クリークを近似的に探索する手法を提案した [12] (以降 RndMCP 法と呼ぶ)。ExMCP 法 [11] の問題点はテスト構成条件を満たす頂点が爆発的に増加し、グラフの探索空間が莫大となることである。そのため、RndMCP 法ではテスト構成候補グラフから部分グラフをランダムに抽出し、ここから最大クリーク探索を繰り返すことにより、グラフ全体の最大クリークを全体の最大クリークを近似的に探索する。

この手法は L_1 個の頂点からなる部分グラフにおける最大クリーク探索を時間 L_2 だけ行うため、時間・空間計算量をそれぞれ $O(L_2)$ 、 $O(L_1^2)$ に緩和する。また、最大クリーク探索は時間 CT だけ繰り返す。これらのパラメータは計算機環境に合わせて任意に設定できるため、ExMCP 法 [11] の時間・空間計算量 $O(2^{|V|})$ 、 $O(|V|^2)$ に対して格段に扱いやすい。

この手法により一般的な規模 (500~2000 項目程度) のアイテムバンクから最大で 10 万程度のテストを生成できた。

3.2 整数計画法を用いた等質テスト構成

RndMCP 法ではグラフの頂点数を $|V|$ としたとき、最先端の最大クリーク探索アルゴリズム [13], [14] を用いたとしても空間計算量を $O(|V|^2)$ 必要とするため 10

万個程度のテスト構成が上限であった。そこで、Ishii et al. (2017) は RndMCP 法と整数計画法を組み合わせた HybridRBP 法を提案した [15], [16]。

HybridRBP 法では現在探索中のクリーク C の全頂点と隣接する頂点を整数計画法を用いて、逐次的に探索する。本手法では現在探索中のクリークに隣接する頂点のみを保存するため、グラフの頂点数を $|V|$ とすると、計算に必要な空間計算量は $O(|V|)$ に軽減される。ただしこの探索は $O(|V| \cdot 2^n)$ の時間計算量を必要とするため、RndMCP 法の最大クリーク探索の時間計算量 $O(L_2)$ に大幅に劣る。そこで RndMCP 法により計算機のメモリの限界の頂点数 L_1 をもつグラフで最大クリーク探索を行ってから、整数計画法を用いる手法に切り替えることで探索効率を改善する [15], [16]。具体的には次の整数計画問題でクリークに隣接する頂点を探索する。

最大クリーク探索のための整数計画問題

variables

$$x_i = \begin{cases} 1 & i \text{ 番目の項目がテストに含まれる} \\ 0 & \text{それ以外} \end{cases} \quad (4)$$

maximize

$$\sum_{i=1}^n \lambda_i x_i \quad (5)$$

subject to

$$\sum_{i=1}^n x_i = M \text{ (テスト項目数)} \quad (6)$$

$$LB_{\theta_k} \leq \sum_{i=1}^n I_i(\theta_k) x_i \leq UB_{\theta_k} \quad (7)$$

$$(k = 1, \dots, K)$$

$$\sum_{i=1}^n X_{i,r} x_i \leq OC \text{ (項目重複上限数)} \quad (8)$$

$$(r = 1, \dots, |C|)$$

$$X_{i,r} = \begin{cases} 1 & i \text{ 番目の項目が} \\ & C \text{ 中の } r \text{ 番目のテストに含まれる} \\ 0 & \text{それ以外} \end{cases}$$

制約条件の式 (8) はクリーク C の全頂点と隣接するための条件である。

目的関数に含まれる $\lambda_i (i = 1, 2, \dots, n)$ は互いに独立な $[0, 1]$ の連続一様分布であり、本問題が解かれるごとにリサンプリングされ毎回ランダムにテストが構成される。この定式化は Belov (2008) [6] で用いられたランダムにテスト構成を行う整数計画問題への定式化を項目重複について一般化したものとなっている [16].

この手法の空間計算量は内部で使用する整数計画法の空間計算量が無視できるとすると、 $O(|V|)$ となる。RndMCP 法と比較して空間計算量が $O(|V|^2)$ から $O(|V|)$ に減少しているため、構成可能なテストの上限は大きくなる。これにより、乱択法では 10 万個のテスト構成が上限であったが、HybridRBP 法では 10 万を超えるテスト構成が可能になった。

3.3 露出率を軽減するテスト構成手法

先行研究 [12], [15], [16] では、ある項目がテスト構成全体で出題される回数（露出数）に偏りが生じる問題がある。露出数が大きい項目は受験者間で共有されやすく、その項目の信頼性が失われやすくなる [18]. ここで、等質テスト群（クリーク C ）に含まれる項目 i の露出数 IEC_i は以下のように表される。

$$IEC_i = \sum_{r=1}^{|C|} X_{i,r} \quad (9)$$

$$X_{i,r} = \begin{cases} 1 & (\text{項目 } i \text{ がテスト } r \text{ に含まれる}) \\ 0 & (\text{項目 } i \text{ がテスト } r \text{ に含まれない}) \end{cases}$$

また、テスト構成 C における最大露出数を

$$IEC_{\max} = \max_{i=1 \dots n} IEC_i \quad (10)$$

で表す。このときテスト構成 C における項目露出率を

$$IEC_{\text{rate}} = \frac{IEC_{\max}}{|C|} \quad (11)$$

とする。

Ishii and Ueno (2015) はこの露出率を軽減するテスト構成を提案した [19]. 本手法は RndMCP 法 (3.2) と整数計画問題を用いた最大クリーク探索 (3.3) を用いてテスト構成し、そのテスト構成の中で露出率が最小となるようなテスト構成を選択する。

具体的にはアルゴリズム 1 により、テスト構成を行う。以降、この手法を Maximum Clique Algorithm with Low Item Exposure (MCALIE) 法と呼ぶ。

MCALIE 法は大きく分けて “initialize”, “add step”, “delete step”, “output” から構成される。

Algorithm 1 MCALIE 法

Require: アイテムバンク、テスト構成条件

Ensure: 等質テスト群

```

1: procedure MCALIE( $L_1, L_2, CT', \alpha, CT$ )
2:    $ST := \text{current time}$ 
3:   /* initialize */
4:   global  $C := \text{RndMCP}(L_1, L_2, CT')$ 
5:   global  $C_{\text{candidate}} := \emptyset$ 
6:   while ( $\text{current time} - ST$ ) <  $CT$  do
7:     /* add step */
8:      $Sol := \text{IPSolve}(\text{itemBank}, C)$  ▶ 式 (5)～式 (8) を解く
9:     if  $Sol \neq \emptyset$  then ▶ IP が解けた場合
10:        $C := C \cup \{Sol\}$ 
11:        $C_{\text{candidate}} := C_{\text{candidate}} \cup \{C\}$ 
12:     else ▶ IP が解けない場合
13:       DeleteStep( $|C|, \alpha$ )
14:     end if
15:   end while
16:   /* output */
17:    $C_{\text{res}} := C$ 
18:   for  $C_{\text{cand}}$  in  $C_{\text{candidate}}$  do
19:     if  $\frac{E_{C_{\text{res}}}}{|C_{\text{res}}|} \leq \frac{E_{C_{\text{cand}}}}{|C_{\text{cand}}|}$  then
20:        $C_{\text{res}} := C_{\text{cand}}$  ▶ 露出率が最小の等質テストを選択
21:   end if
22: end for
23: return  $C_{\text{res}}$ 
24: end procedure
25: procedure DeleteStep( $\text{cliqueSize}, \alpha$ )
26:   /* delete step */
27:    $\text{count} := 0$ 
28:   while  $\text{count} < (\text{cliqueSize} \times \alpha)$  do
29:      $C := C \setminus \{c \in C\}$  ▶ 最大露出項目をもつテストから除外する
30:      $\text{count}++$ 
31:   end while
32:    $C_{\text{candidate}} := C_{\text{candidate}} \cup \{C\}$ 
33: end procedure

```

“initialize” では RndMCP 法によりメモリ限界 L_1 の制限下で最大クリーク C を探索する。

“add step” では整数計画法により最大クリーク C に隣接する頂点を逐次的に探索し C に追加することで、より大きなテストを生成する。生成したテスト C は $C_{\text{candidate}}$ に保存する。

“delete step” では “add step” で整数計画法が解けなかった場合、 C から一定割合 (α) だけ頂点を除外する。この際、最大露出の項目を含むテストから削除することで露出数を減らしている。削除によって変更されたクリーク C は $C_{\text{candidate}}$ に保存する。

“output” では生成したテストの集合 $C_{\text{candidate}}$ から露出率 $\frac{E_{C_{\text{res}}}}{|C|}$ が最も低いテスト C を選択し出力する。

これらの提案により、従来手法 [4], [7], [8], [12] よりも露出率が軽減されている [19].

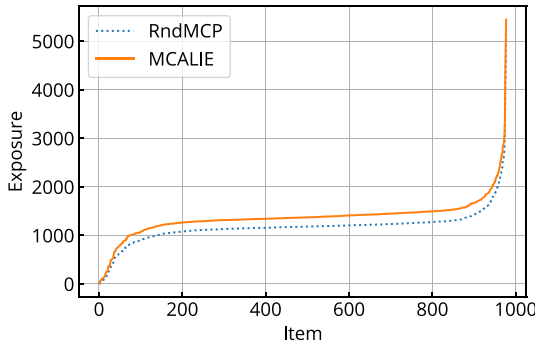


図4 MCALIE法の露出数の分布
Fig. 4 Distribution of exposure with MCALIE.

表2 MCALIE法の再現実験の結果
Table 2 The results of MCALIE experiment.

Pool Size	OC	RndMCP			MCALIE		
		$ C $	IEC_{max}	IEC_{rate}	$ C $	IEC_{max}	IEC_{rate}
978 (actual)	0	18	1	5.6%	36	1	2.8%
	5	45790	5177	11.3%	51644	5342	10.3%
	10	100000	16310	16.3%	102862	16698	16.2%

3.4 MCALIE法の問題

MCALIE法[19]は構成した等質テスト群を保存しておき、最後にその中から最も露出率が低いテスト群を選択することで露出率を軽減している。しかし、等質テストを構成する際に露出を考慮していないため、特定の項目が過度に露出（出題）される問題を解消できていない。

この従来手法の問題を明らかにするために、MCALIE法の再現実験を行った。ここで、再現実験は978項目からなる実際に運用されたアイテムバンクを用い、テスト項目数を25、重複項目数 $OC=\{0, 5, 10\}$ として実験を行った。パラメータ値はRndMCP法が $L_1 = 100000, L_2 = 4hr, CT = 4hr$ 、MCALIE法が $L_1 = 100000, L_2 = 4hr, CT' = 4hr, CT = 24hr$ である。

結果を表2に示す。MCALIE法はRndMCPよりもテスト構成数が増大し、露出率は減少はしたものの、依然として露出率は大きい。図4は $OC = 5$ の再現実験で構成されたテストの項目露出数を昇順に並べ替えたものである。具体的には縦軸に露出数、横軸は露出数の低い順に並べ替えられた項目を示している。図4のように一部項目は露出数が非常に大きく（グラフの右端）、テスト全体の10%以上に出題されている。MCALIE法も特定の項目が優先的に選択され、過度な露出を引き起こしていることが分かる。

4. 提案手法

MCALIE法[19]の第1段階目のRndMCP法はランダム探索による最大クリーク探索であり、頂点集合における露出数の偏りが生じる問題がある。この問題を解決するために、本論文では整数計画法を用いて頂点生成する。提案手法の整数計画法ではテストの構成条件を満たす項目集合のうち最も目的関数（式(5)）の値が大きいものを選択するが、この目的関数の係数 λ_i は一樣分布に従う乱数であるため、制約を満たす頂点集合から頂点が一樣に生成される。このため、第1段階目の頂点生成において項目露出の偏りを改善することができる。

しかしこの整数計画法は項目ごとの露出数の一樣性は考慮していないため、依然として特定の項目が過度に出題されてしまうバイアス問題がある。この問題を解決するために、頂点集合における露出数上位 s 位までの項目をテスト構成ごとにアイテムバンクから除外し、テスト構成を行う手法を提案する。これにより、露出数上位の項目を除いたアイテムバンクからテスト構成するので、最大露出数が抑制され、テスト構成の項目露出数が一樣になることが期待される。

具体的には、第1段階目の頂点生成において項目重複数条件を除いた整数計画問題（式(5)～式(7)）を解き、頂点生成し頂点集合に追加する。頂点生成後に、全ての項目のうち頂点集合における露出数上位 s 位までの項目をアイテムバンクから除外する。以上を頂点集合の大きさが L_1 になるまで逐次的に繰り返し頂点を生成する。ただし、新たに頂点を生成できない場合は今までに除外した項目を全てアイテムバンクに戻す。その後、RndMCP法と同様に頂点集合から構築されるグラフの最大クリークを探索する。

提案手法では露出数の大きい項目が除外されたアイテムバンクからテスト構成するため、図4のように特定の項目が選択され続けることを防ぎ、項目がテスト群に均等に追加され、従来手法よりも露出数の偏りを軽減できると期待される。

本研究では、グラフの頂点集合（テスト集合）を V とし、式(12)の項目 i の V における露出数 IEV_i を導入する。

$$IEV_i = \sum_{r=1}^{|V|} X_{i,r} \quad (12)$$

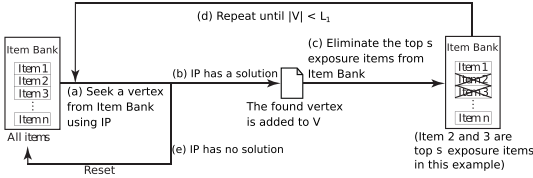


図5 第1段階目の提案手法

Fig. 5 The proposed method for the first step.

$$X_{i,r} = \begin{cases} 1 & \text{(項目 } i \text{ がテスト } r \text{ に含まれる)} \\ 0 & \text{(項目 } i \text{ がテスト } r \text{ に含まれない)} \end{cases}$$

提案手法は概要図5のように、 IEV_i を用いて以下の手順に従い頂点を生成する。

- (a) 整数計画問題（式(5)～式(7)）を解く。
- (b) 整数計画問題が解けた場合は、その解を V に追加する。
- (c) 露出数 IEV が上位 s 位までの全ての項目をアイテムバンクから除外する。（図5の中では項目2と3が該当する。）
- (d) (a)～(c)の手順を L_1 個の頂点を生成するまで繰り返す。
- (e) 整数計画問題が解けない場合は、除外した項目を全てアイテムバンクに戻す。

整数計画問題はNP困難であるため、一定の計算時間IPTで解を得られない場合は問題が解けないとみなし、(e)で今までに除外した全ての項目をアイテムバンクに戻す。

頂点生成後はRndMCPと同様にして項目重複数条件を満たす2頂点間に辺を引いてグラフを構築し、そのグラフにおいて、最大クリーク探索を行う。

この提案手法を組み込んだRndMCP法のアルゴリズムをAlgorithm2に示す。

step1では最大クリーク探索するための部分グラフの頂点を L_1 個生成する。その頂点集合を V とする。 $allowedItems$ は IEV が上位 s 位までの項目をアイテムバンクから除いた項目集合である。項目重複数条件（式(8)）を除いた整数計画問題（式(5)～式(7)）を解き、頂点(Sol)を生成する。整数計画問題の解を求めることができない場合は、 $allowedItems$ に除外した項目を戻し、再度整数計画問題を解く。頂点に含まれる

Algorithm 2 選択可能な項目を制限した RndMCP 法

Require: アイテムバンク、テスト構成条件

Ensure: 等質テスト群

```

1: procedure LimitItemRndMCP( $L_1, L_2, CT, s$ )
2:    $C := 0, C_{\max} := 0$ 
3:    $ST := \text{current time}$ 
4:   while ( $\text{current time} - ST$ ) <  $CT$  do
5:     /* Step1 */
6:      $V := \emptyset$ 
7:      $n := \text{itemBank.size}$ 
8:     for  $i \leftarrow 0$  to  $n - 1$  do
9:        $IEV[i] := 0$ 
10:    end for
11:     $allowedItems := \text{アイテムバンクの全項目}$ 
12:    while  $|V| < L_1$  do
13:       $Sol := \text{IPSolve}(allowedItems)$ 
14:       $\triangleright allowedItems$  のみからなるアイテムバンクに対し
        式(5)～式(7)の整数計画問題を解く
15:      if  $Sol = \emptyset$  then  $\triangleright$  整数計画問題が解けない場合
16:         $allowedItems := \text{アイテムバンクの全項目}$ 
17:        break
18:      end if
19:       $V := V \cup \{Sol\}$ 
20:      for each  $i \in Sol$  do
21:         $IEV[i] ++$ 
22:      end for
23:       $allowedItems := IEV$  の上位  $s$  位以外の項目
24:    end while
25:    /* Step2 */
26:     $G = (V, E)$  グラフ構築  $\triangleright$  二頂点が重複項目数条件を満たす場合、辺を引く
27:    /* Step3 */
28:     $C := \text{MCP}(G, L_2)$   $\triangleright G$  の最大クリークを時間  $L_2$  だけ探索
29:    if  $|C_{\max}| < |C|$  then
30:       $C_{\max} := C$ 
31:    end if
32:  end while
33:  return  $C_{\max}$ 
34: end procedure

```

項目について IEV をインクリメントする。

step2ではstep1で生成した頂点集合 V からグラフ G を構築する。具体的には項目重複数条件を満たす2頂点間に辺を引きグラフを構築する。

step3ではstep2で構築したグラフ G において最大クリーク探索をして、等質テスト構成をする。

5. 評価実験

提案手法の有効性を示すために、評価実験を行う。最初に、RndMCP法と提案手法の整数計画法を比較する。次に、提案手法の露出数上位 s 位までの項目を除外するアルゴリズムのパラメータの最適化及び選択項目の特性の分析を行う。最後に、第2段階目でMCALIE法とHybridRBPの第1段階目の手法としてRndMCP法

表3 アイテムバンクのパラメータ分布の詳細

Table 3 Details of the parameter distribution of the item bank.

Pool Size	parameter a		parameter b	
	Mean	SD	Mean	SD
500	0.49	0.28	0.05	1.06
1000	0.49	0.28	0.01	1.00
2000	0.51	0.29	0.04	1.00
978(actual)	0.43	0.20	-0.22	1.16

表4 実験で用いたテスト情報量の上下限制約

Table 4 Test information constraint for the experiment.

$\theta = -2.0$	$\theta = -1.0$	$\theta = -0.0$	$\theta = 1.0$	$\theta = 2.0$
2.0/2.4	3.2/3.6	3.2/3.6	3.2/3.6	2.0/2.4

と提案手法を用いたものとで比較し、提案手法の有効性を示す。なお、本論文の実行環境は Ubuntu18.04 を OS とする計算機 (CPU: Intel Core i9-9900X 3.50 GHz, RAM 128 GB) である。

本論文の実験では、実際に運用されていた実アイテムバンク及びシミュレーションアイテムバンクを用いた。実アイテムバンクは 978 項目、シミュレーションアイテムバンクは 500, 1000, 2000 項目をもち、項目の識別力パラメータ a を $\log_2 a \sim N(0, 1^2)$ 、困難度パラメータ b を $b \sim N(0, 1^2)$ として実アイテムバンクのパラメータ値の分布に近づけるように発生させた。生成したシミュレーションアイテムバンク及び実アイテムバンクのパラメータ値の分布の詳細は表3のとおりである。

テストの構成条件は前述したアイテムバンクから、表4のテスト情報量の上下限制約を満たす 25 項目のテスト構成とした。本条件は実際に運用された e テスティングにおけるテスト構成条件である。

項目重複数条件は先行研究 [19] と同様に $OC = \{0, 5, 10\}$ の 3 通りの条件によって評価する。なお、従来手法 [19] 及び提案手法の整数計画法の求解には CPLEX12.9 [26] を用い、整数計画緩和の解との相対ギャップが 10^{-4} 以下で打ち切るようにした (デフォルトのオプション)。更に整数計画問題の計算時間の上限を $IPT = 60\text{sec}$ と設定した。

5.1 RndMCP 法と整数計画法の比較実験

第 1 段階目の頂点生成において、提案手法の整数計画法を用いることで露出数の偏りが改善されることを示すために、RndMCP 法 [12] と比較実験を行った。

パラメータ値は $L_1 = 100000$, $L_2 = 4\text{hr}$, $CT = 4\text{hr}$ とした。ただし、 $n = 978$ の場合は計算時間 $CT = 4\text{hr}$ 以内に頂点を L_1 個生成できなかったため、 $CT = 8\text{hr}$

表5 RndMCP 法と整数計画法の比較実験の結果

Table 5 The results of RndMCP and IP comparison experiments.

PoolSize	OC	RndMCP				提案手法 (整数計画法)			
		$ C $	IEC_{σ}	$IEC_{\max}$	IEC_{rate}	$ C $	IEC_{σ}	$IEC_{\max}$	IEC_{rate}
500	0	10	0.5	1	10.0%	10	0.5	1	10.0%
	5	4371	53.1	378	8.6%	4952	50.5	402	8.1%
	10	99981	1729.2	13006	13.0%	99979	1153.2	7259	7.3%
1000	0	17	0.5	1	5.9%	18	0.5	1	5.6%
	5	46190	430.5	3374	7.3%	52023	389.1	2206	4.2%
	10	100000	990.7	8767	8.8%	100000	755.4	4139	4.1%
2000	0	32	0.5	1	3.1%	31	0.5	1	3.2%
	5	96773	456.1	3833	4.0%	88651	325.2	1674	1.9%
	10	100000	472.6	4044	4.0%	90917	333.7	1707	1.9%
978 (actual)	0	18	0.5	1	5.6%	19	0.5	1	5.3%
	5	45790	368.2	5177	11.3%	54964	251.7	2303	4.2%
	10	100000	985.3	16310	16.3%	100000	488.2	4301	4.3%

と計算時間を延長した。

ここで、各項目の露出数 IEC_i の平均を

$$IEC_{\mu} = \frac{\sum_{i=1}^n IEC_i}{n} \quad (13)$$

とし、各項目の露出数 IEC_i の標準偏差を

$$IEC_{\sigma} = \sqrt{\frac{\sum_{i=1}^n (IEC_i - IEC_{\mu})^2}{n}} \quad (14)$$

とする。

重複項目数を OC 、テスト構成数を $|C|$ 、最大露出数を $IEC_{\max}$ 、露出率を IEC_{rate} として、露出数の標準偏差 IEC_{σ} とともに結果を表5に示す。なお、露出数の平均値 IEC_{μ} については手法間でテスト構成数が異なると、単純に比較することができないため表には載せていない。

$OC = 0$ 以外の条件下で整数計画法の方が露出数の標準偏差及び露出率が小さくなった。提案手法の整数計画法では制約を満たす頂点集合から頂点が一樣に選択され、整数計画法を用いることにより露出率が改善されることが示された。

図6は $n = 978$, $OC = 5$ における RndMCP 法と提案手法 (整数計画法 (IP)) によるテスト構成の露出数の分布である。縦軸は露出数、横軸は項目を表し、項目は露出数の昇順に並び替えている。図6からも整数計画法により露出数の偏りが改善されたことが確認できる。しかし、特定の項目に偏って露出される問題は依然として存在する。図6の整数計画法では特定の項目 (グラフの右端) が 2303 個のテストに含まれ、テスト全体の 4.2% に出題される。本提案手法の整数計画問題により制約を満たす頂点集合から頂点を一樣に生成できるが、各項目ごとの露出数の一様性については考慮していないので、依然として特定の項目に偏って露出してしまふ。

$n = 2000$ 以外の $OC = 5$ の条件では、テスト構成数

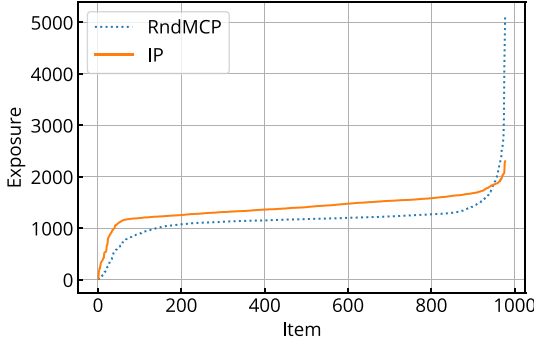


図6 RndMCP 法と整数計画法の露出数の分布

Fig. 6 Exposure Distribution of RndMCP and Integer Programming.

が RndMCP 法よりも大きくなった。そこで、テスト構成数の増加の原因について分析するために、最大クリーク探索を行うグラフの頂点集合 V における露出数 IEV_i 式 (9) と見つかった最大クリーク C における露出数 IEC_i 式 (12) を用いて、項目 i を含む頂点のうち、最大クリークに選ばれた割合 CR_i を以下のように定義する。

$$CR_i = \frac{IEC_i}{IEV_i} \quad (15)$$

この CR_i が大きいほど、項目 i を含む頂点が最大クリーク C に選ばれる割合が高くなる。

図 7 は $n = 978, OC = 5$ の条件下で横軸に露出数 IEV_i 、縦軸に CR_i をとったグラフである。露出数 IEV_i が大きいほど、 CR_i が小さくなる。これは露出数が高い項目を含む頂点は最大クリークに選ばれる割合が低くなることを示している。したがって、整数計画法を用いて、露出数 IEV_i を抑えたグラフから探索を行うことで、最大クリークに選ばれる頂点の割合を高め、テスト構成数を増加させたと考えられる。また、RndMCP 法と提案手法を比較すると、露出数 IEV_i が同じ値でも CR_i が大きい。これは提案手法の露出数 IEV_i の最大値が小さくなることで、最大クリークに選ばれる頂点の割合が全体的に高くなることを示している。

比較実験により、第 1 段階目のテスト構成において整数計画法を用いることで、露出率や露出数の偏りを抑えるだけでなく、テスト構成数も増加する場合があることが分かった。この提案手法を先行研究 [16], [17] の第 1 段階目に用いることで、テスト構成数が更に増加することが期待できる。

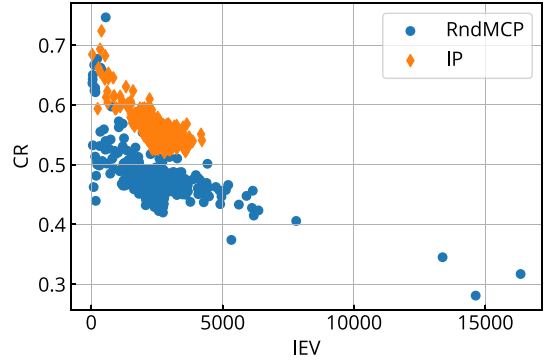


図7 露出数とクリークの関係

Fig. 7 Relation between exposure and clique.

5.2 提案手法における s の最適化

前節のとおり、本提案手法の整数計画法では項目露出数の一様性が考慮されておらず、特定の項目が過度に出題されてしまう。この問題を解決するために、提案手法では露出数が上位 s 位までの項目をアイテムバンクから除外し、残りの項目からテストを生成する。本章ではこのパラメータ値 s を変化させたときの露出率を比較し、露出率が最小となる s の最適値を決定する。具体的には $s = \{0, 1, 2, 3\}$ の 4 通りの値で露出率を比較する。なお、 $s = 0$ の場合はアイテムバンクからどの項目も除外しないことを意味し、前節 (5.1) の整数計画法と同等である。

パラメータ値は $L_1 = 100000$, $L_2 = 4hr$, $s = \{0, 1, 2, 3\}$, $CT = 4hr$ とした。ただし、 $n = 978$ の場合は計算時間 $CT = 4hr$ 以内に頂点を L_1 個生成できなかったため、 $CT = 8hr$ と計算時間を延長した。

結果を表 6 に示す。アイテムバンクから項目を除外しない $s = 0$ の場合と項目を除外する $s \geq 1$ の場合を比較すると、全ての n, OC の条件において、 $s \geq 1$ の場合の方が露出率が小さい。これは $s = 0$ で整数計画法でテストを構成するよりも、 $s \geq 1$ で項目を除外してテストを構成することで特定の項目が過度に出題されないことを示している。更に、 $s \geq 1$ の露出率を比較すると、 $n = 500$, $OC = 5$ の条件を除いて $s = 1$ の場合が最も露出率が小さい。また、表 6 で s の値が 1 以上の場合には露出率が大きく変化していない。このことは s が 1 のときに広い範囲で一様に露出数の高い項目を除外できていることを示し、 $s=1$ が有効に機能していることが分かる。図 8 は $n = 978$, $OC = 5$ の条件下で RndMCP 法と露出数上位 s 位までの項目を除外する提案手法を用いた場合のテスト構成の露出数の

表 6 s の最適化実験の結果Table 6 The results of s optimization experiments.

Pool Size	OC	$s = 0$			$s = 1$			$s = 2$			$s = 3$		
		$ C $	$IEC_{\max}$	IEC_{rate}	$ C $	$IEC_{\max}$	IEC_{rate}	$ C $	$IEC_{\max}$	IEC_{rate}	$ C $	$IEC_{\max}$	IEC_{rate}
500	0	10	1	10.0%	11	1	9.1%	10	1	10.0%	11	1	9.1%
	5	4952	402	8.1%	4994	379	7.6%	5015	375	7.5%	4936	373	7.6%
	10	99979	7259	7.3%	99970	5370	5.4%	99955	5391	5.4%	99946	5428	5.4%
1000	0	18	1	5.6%	18	1	5.6%	18	1	5.6%	18	1	5.6%
	5	52023	2206	4.2%	50779	1616	3.2%	50645	1674	3.3%	50175	1653	3.3%
	10	100000	4139	4.1%	99998	2847	2.8%	99998	2859	2.9%	99998	2888	2.9%
2000	0	31	1	3.2%	32	1	3.1%	32	1	3.1%	32	1	3.1%
	5	88651	1674	1.9%	97320	1397	1.4%	97380	1398	1.4%	97251	1399	1.4%
	10	90917	1707	1.9%	100000	1418	1.4%	100000	1419	1.4%	100000	1421	1.4%
978 (actual)	0	19	1	5.3%	19	1	5.3%	19	1	5.3%	19	1	5.3%
	5	54964	2303	4.2%	54800	1683	3.1%	54853	1695	3.1%	54565	1677	3.1%
	10	100000	4301	4.3%	100000	2754	2.8%	100000	2760	2.8%	99998	2770	2.8%

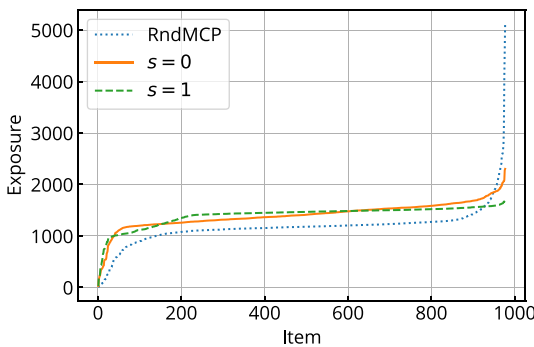


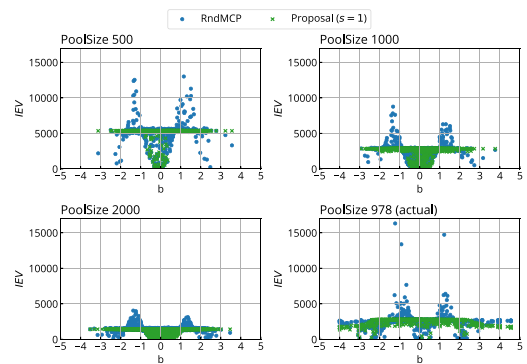
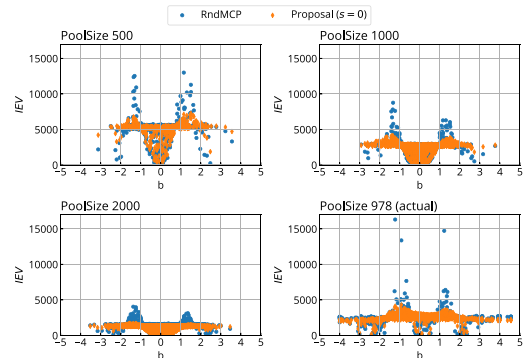
図 8 RndMCP 法と提案手法の露出数の分布

Fig. 8 Exposure distribution of RndMCP and proposed method.

分布である。縦軸は露出数、横軸は項目を表し、項目は露出数の昇順に並び替えている。図 8 から、露出数上位 1 位の項目を除外することで特定の項目（右端）が過度に出題される問題が改善されることが確認できる。以降の実験では $n = 500$, $OC = 5$ の条件で $s = 2$, それ以外の条件では $s = 1$ を提案手法の s の最適値とする。

5.3 RndMCP 法と提案手法により選択される項目特性の分析

提案手法により特定の項目が過度に出題される問題が改善された。そこで、RndMCP 法と従来手法で選択される項目特性の差異を分析する。図 9 は $n = 500, 1000, 2000, 978$ (実アイテムバンク), $OC = 10$ の条件下で RndMCP 法と提案手法 ($s = 1$) により選択された項目のパラメータ b を横軸に、露出数 IEV を縦軸にプロットした図である。図 10 は図 9 と同様の条件で RndMCP 法と提案手法 ($s = 0$) により選択された項目パラメータ b と露出数をプロットした

図 9 RndMCP 法と提案手法 ($s = 1$) による項目パラメータ b と露出数 IEV の関係Fig. 9 Relation of item parameter b , and exposure IEV between RndMCP and proposed method ($s = 1$).図 10 RndMCP 法と提案手法 ($s = 0$) による項目パラメータ b と露出数 IEV の関係Fig. 10 Relation of item parameter b , and exposure IEV between RndMCP and proposed method ($s = 0$).

図である。ただしいずれの図においても $n = 978$ のアイテムバンクにおいて一度も出題されなかった項目 ($a = 3.08, b = -0.11$) は除外してプロットした。

$n = 500, 1000, 978$ のアイテムバンクで、 $b \in (-2, -1)$, $b \in (1, 2)$ の項目が RndMCP 法や提案手法 ($s = 0$) では過度に選択されていることが分かる。 b の分布は標準正規分布に従うため $b \in (-2, -1)$, $b \in (1, 2)$ の項目数は $b \in (-1, 1)$ と比較して少ないが、テスト構成では能力の高い (低い) 少数の受験者の測定精度も保証するため、RndMCP 法は露出数を考慮せず、これらの能力に対応する少数項目を重複して選択していることが理由と考えられる。一方で、 $n = 2000$ の比較的大きいアイテムバンクでは露出数の偏りが小さい (図 9, 図 10)。これは $n = 2000$ では $b \in (-2, -1)$, $b \in (1, 2)$ の項目数が多いため、露出数を考慮していない RndMCP 法や提案手法の $s = 0$ のアルゴリズムでも重複せずに項目選択でき、露出数の偏りが小さくなるからと考えられる。図 11 は b の範囲を項目露出数の大きい $b \in (-2, -1)$, $b \in (1, 2)$ とそれ以外 (図中では「otherwise」) に分けて、パラメータ a を横軸に、露出数 IEV を縦軸にプロットした図である。 a については RndMCP 法では、 a の値が大きくなるにつれて $b \in (-2, -1)$, $b \in (1, 2)$ の項目の露出数は大きくなり、それ以外の b の値の項目の露出数は小さくなる (図 11)。これは、図 9 の分析で示したように b の範囲に対する項目数の差異が影響していると考えられる。一方で、提案手法は露出数を考慮して項目を選択するため、特定の b や a 付近の値ををもつ少数の項目が過度に出題されるという特性を抑制できている。

次に、提案手法のアルゴリズムにより頂点生成時に除外される項目の特性について分析する。図 12 はアイテムバンク $n = 500, 1000, 2000, 978, OC = 10$ の条件下で提案手法 ($s = 1$) で頂点生成した際の項目のパラメータ b を横軸に、除外した項目の回数を縦軸にプロットした図である。図 12 より $n = 500, 1000, 978$ のアイテムバンクでは、提案手法が $b \in (-2, -1)$, $b \in (1, 2)$ の項目を多く除外して頂点を生成したことが分かる。これは図 9, 図 10 での分析で示したように $n = 500, 1000, 978$ では $b \in (-2, -1)$, $b \in (1, 2)$ の項目数が少ないため、頂点生成のたびに重複して選択されやすく露出数が上位になるので、多く除外されるからである。一方で、 $n = 2000$ では $b \in (-2, -1)$, $b \in (1, 2)$ の項目数が大きく、これらの項目は重複して選択されずに露出数の偏りが生じにくい。したがって、アルゴリズムにより $b \in (-2, -1)$, $b \in (1, 2)$ だけではなく $b \in (-1, 1)$ の項目も除外されていることが分かる (図 12)。以上から、露出数上位 1 位までの項目を除外する提案手法 ($s = 1$)

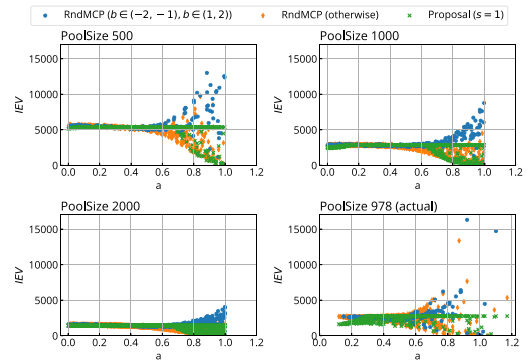


図 11 RndMCP 法と提案手法による項目パラメータ a と露出数 IEV の関係

Fig. 11 Relation of item parameter a , and exposure IEV between RndMCP and proposed method.

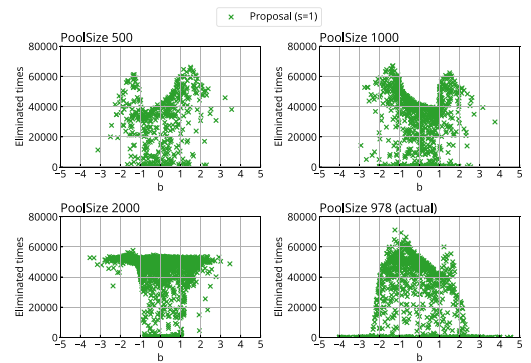


図 12 提案手法により除外された項目パラメータ b の分布

Fig. 12 Distribution of Item parameter b and number that were eliminated by the proposal method.

はアイテムバンクサイズが大きい場合に露出数の軽減に寄与していることが分かる。

5.4 第 1 段階目と第 2 段階目を統合したテスト構成の比較実験

第 1 段階目で提案手法を用いることによる第 2 段階目のテスト構成のテスト構成数や露出率への影響を確認するために、第 1 段階目で RndMCP 法 [12] を用いた場合と提案手法を用いた場合で第 2 段階目の比較実験を行った。第 2 段階目のテスト構成手法は MCALIE 法 [19] と HybridRBP 法 [15], [16] でテスト構成数及び露出率を比較した。各手法の計算時間は 24hr とし、HybridRBP 法と MCALIE 法のパラメータは $L_1 = 100,000$, $L_2 = 4\text{hr}$, $\alpha = 0.1$, $CT' = 4\text{hr}$ (シミュレーションデータ), $CT' = 8\text{hr}$ (実データ), $CT = 24\text{hr}$ とした。第 1 段階目の提案手法のテスト構成は前述の比較実験 (5.2) で最も露出率が小さいものを採用した。

表 7 第 1 段階目と第 2 段階目を統合したテスト構成の実験結果

Table 7 The results of first step and second step test assembly experiments.

Pool Size	OC	RndMCP/MCALIE				RndMCP/HybrdRBP				提案手法/MCALIE				提案手法/HybrdRBP			
		C	IEC _σ	IEC _{max}	IEC _{rate}	C	IEC _σ	IEC _{max}	IEC _{rate}	C	IEC _σ	IEC _{max}	IEC _{rate}	C	IEC _σ	IEC _{max}	IEC _{rate}
500	0	18	0.3	1	5.6%	18	0.3	1	5.6%	18	0.3	1	5.6%	17	0.4	1	5.9%
	5	11475	130.8	858	7.5%	15795	177.1	1216	7.7%	13430	143.3	998	7.4%	16016	171.2	1192	7.4%
	10	104914	1778.0	13267	12.6%	104817	1777.8	13332	12.7%	99970	1138.4	5370	5.4%	104684	1187.0	5720	5.5%
1000	0	36	0.3	1	2.8%	35	0.3	1	2.9%	35	0.3	1	2.9%	34	0.4	1	2.9%
	5	55047	492.1	3691	6.7%	55087	492.5	3712	6.7%	50781	377.7	1616	3.2%	59316	438.9	1986	3.3%
	10	104260	1019.9	8762	8.4%	104228	1019.5	8724	8.4%	99998	769.5	2847	2.8%	104250	799.2	3031	2.9%
2000	0	70	0.3	1	1.4%	69	0.3	1	1.4%	70	0.3	1	1.4%	70	0.3	1	1.4%
	5	99620	466.5	4028	4.0%	99783	467.0	4031	4.0%	97348	378.0	1397	1.4%	100196	387.9	1452	1.4%
	10	102775	482.4	4105	4.0%	102849	482.6	4106	4.0%	100000	388.7	1418	1.4%	102839	398.5	1480	1.4%
978(actual)	0	36	0.3	1	2.8%	35	0.3	1	2.9%	36	0.3	1	2.8%	35	0.3	1	2.9%
	5	51644	388.7	5342	10.3%	51984	389.8	5357	10.3%	54800	204.6	1683	3.1%	59668	223.4	1868	3.1%
	10	102862	996.1	16698	16.2%	102624	995.7	16686	16.3%	100000	418.7	2754	2.8%	102765	428.0	2881	2.8%

表 8 第 1 段階目のテスト構成の実験結果

Table 8 The results of first step test assembly experiments.

Pool Size	OC	RndMCP				提案手法			
		C	IEC _σ	IEC _{max}	IEC _{rate}	C	IEC _σ	IEC _{max}	IEC _{rate}
500	0	10	0.5	1	10.0%	11	0.5	1	9.1%
	5	4371	53.1	378	8.6%	5015	50.7	375	7.5%
	10	99981	1729.2	13006	13.0%	99970	1138.4	5370	5.4%
1000	0	17	0.5	1	5.9%	18	0.5	1	5.6%
	5	46190	430.5	3374	7.3%	50779	377.7	1616	3.2%
	10	100000	990.7	8767	8.8%	99998	769.5	2847	2.8%
2000	0	32	0.5	1	3.1%	32	0.5	1	3.1%
	5	96773	456.1	3833	4.0%	97320	377.9	1397	1.4%
	10	100000	472.6	4044	4.0%	100000	388.7	1418	1.4%
978(actual)	0	18	0.5	1	5.6%	19	0.5	1	5.3%
	5	45790	368.2	5177	11.3%	54800	204.6	1683	3.1%
	10	100000	985.3	16310	16.3%	100000	418.7	2754	2.8%

表 9 第 2 段階目のテスト構成数

Table 9 The number of tests in second step.

Pool Size	OC	RndMCP/MCALIE	RndMCP/HybrdRBP	提案手法/MCALIE	提案手法/HybrdRBP
		C	C	C	C
500	0	8	8	7	6
	5	7104	11424	8415	11001
	10	4933	4836	0	4714
1000	0	19	18	17	16
	5	8857	8897	2	8537
	10	4260	4228	0	4252
2000	0	38	37	38	38
	5	2847	3010	28	2876
	10	2775	2849	0	2839
978(actual)	0	18	17	17	16
	5	5854	6194	0	4868
	10	2862	2624	0	2765

結果を表 7 に示す。第 1 行では第 1 段階目の手法/第 2 段階目の手法として表記している。本論文では第 1 段階目の手法を提案しているの、第 2 段階目に MCALIE 法を用いた場合と HybridRBP 法を用いた場合に分けて分析を行う。

5.4.1 提案手法/MCALIE 法

表 7 より、第 1 段階目に提案手法を用いた MCALIE 法は各手法と比較して多くの条件下で露出率は最小であるが、OC = 0 の場合を除いてテスト構成数は小さ

い。この原因を分析するために表 8 に第 1 段階目のテスト構成の結果を示し、表 9 に第 2 段階目のテスト構成数を示した。表 9 より、第 1 段階目に提案手法を用いた第 2 段階目は MCALIE 法を用いた場合は OC = 0 と n = 500, OC = 5 の場合を除いておおよそテスト構成数が少ないことが分かる。更にこの原因を明らかにするために、図 13 に n = 978, OC = 10 の条件の MCALIE 法におけるテスト構成の露出率の遷移を示す。図 13 より、提案手法/MCALIE 法の組み合わせでは、第 1 段

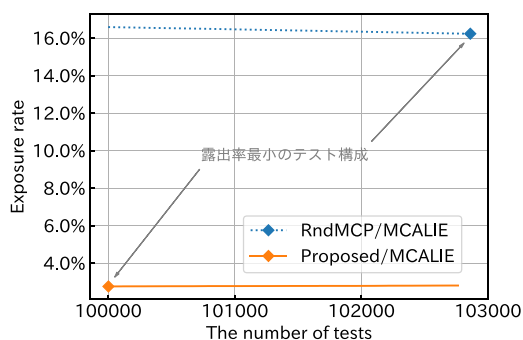


図 13 MCALIE 法における露出率の遷移
Fig. 13 Transaction of exposure rate in MCALIE.

階目で構成したテスト構成が構成した複数のテストの中で露出率が最小であり、MCALIE 法がそのテスト構成を選択し続けるため、テスト構成数が 100,000 から増加しないことが分かる。以上をまとめると、提案手法/MCALIE 法の組み合わせは従来手法と比較して露出率は小さくなるが、MCALIE 法において露出率が第 1 段階目で求めたテスト構成から減少しないため、テスト構成数が少なくなった。

5.4.2 提案手法/HybridRBP 法

表 7 より、第 1 段階目に提案手法を用いた HybridRBP 法は第 1 段階目に RndMCP 法を用いた手法と比較すると $OC = 0$ 以外の全ての条件で、露出率は小さい。テスト構成数は $OC = 5$ では大きく、 $OC = 0, 10$ ではおおそ同様である。 $OC = 5$ でテスト構成数が増加したのは、第 1 段階目の提案手法によるテスト構成数が $OC = 5$ で RndMCP 法よりも増加したからである(表 8)。また表 9 より、 $OC = 5$ では第 2 段階目において第 1 段階目に RndMCP 法を用いたものよりもテスト構成数は少なくなった。これは提案手法では第 1 段階目のテスト構成数が増加したことにより、第 2 段階目の整数計画問題の項目重複数制約(式(8))が多くなり、整数計画問題の求解時間が長くなったからである。以上をまとめると、提案手法/HybridRBP 法の組み合わせは従来手法と比較してテスト構成数を減らすことなく、露出率が低い等質テストを構成できることが分かった。

6. む す び

本論文では e テスティングにおける、項目露出を考慮した整数計画法による等質テスト構成手法を提案した。本手法では RndMCP 法のランダム探索による項

目露出の偏りを整数計画法を用いて制約を満たす頂点集合から頂点を一様に生成することで改善した。しかし、この整数計画法は項目ごとの露出数の一様性は考慮していないため、依然として特定の項目が過度の出題されてしまう問題がある。そこで、露出数が上位の項目を逐次的にアイテムバンクから除外することで、この問題を緩和できた。以上の提案手法の有効性をシミュレーション・実データを用いて示した。

露出率の抑制だけでなく、淵本、植野(2020)の並列アルゴリズム[17]に本手法を組み込むなどして、テスト構成数を大きくするアルゴリズムの開発が今後の課題である。

更に、等質テストは適応型テストに用いることで、テストの長さや項目の露出数を軽減することが知られている。適応型テストとは、受験者の能力を逐次的に推定し、その能力に応じて測定制度が最も高い項目を出題することで、受験時間や項目数を軽減できるコンピュータ・テストの出題方式である。本手法のように、あらかじめ露出数を考慮したテスト群を生成することで、テストの長さや露出数を改善できる可能性がある。このような、実用上の課題についても検討する。

謝辞 本研究の一部は科学研究費補助金(基盤研究(S), 代表: 植野真臣)「信頼性向上を持続する e テスティング・プラットフォームの開発」(19H05663)の助成を受けた。

文 献

- [1] 植野真臣, 永岡慶三, e テスティング, 培風館, 2009.
- [2] M. Ueno, K. Fuchimoto, and E. Tsutumi, "e-testing from artificial intelligence approach," *Behaviormetrika*, vol.48, no.2, pp.409-424, 2021.
- [3] M. Ueno, "AI based e-testing as a common yardstick for measuring human abilities," the 18th International Joint Conference on Computer Science and Software Engineering, pp.1-6, IEEE Comput., 2021.
- [4] W.J. van der Linden, *Liner Models for Optimal Test Design*, Springer, 2005.
- [5] D.I. Belov and R.D. Armstrong, "A constraint programming approach to extract the maximum number of non-overlapping test forms," *Computational Optimization and Applications*, vol.33, no.2-3, pp.319-332, 2006.
- [6] D.I. Belov, "Uniform test assembly," *Psychometrika*, vol.73, no.1, p.21, 2008.
- [7] K.T. Sun, Y.J. Chen, S.Y. Tsai, and C.F. Cheng, "Creating irt-based parallel test forms using the genetic algorithm method," *Applied measurement in education*, vol.21, no.2, pp.141-161, 2008.
- [8] P. Songmuang and M. Ueno, "Bees algorithm for construction of multiple test forms in e-testing," *IEEE Trans. Learning Technolo-*

gies, vol.4, no.3, pp.209–221, 2010.

- [9] P. Songmuang and M. Ueno, “Multiple test forms construction based on bees algorithm,” Educational Data Mining 2010, 3rd Int. Conf. Educational Data Mining, Pittsburgh, PA, USA, June 11–13, 2010. Proceedings, eds. R.S.J. deBaker, A. Merceron, and P.I.P. Jr., pp.321–322, 2010.
- [10] T. Ishii, P. Songmuang, and M. Ueno, “Maximum clique algorithm for uniform test forms assembly,” Artificial Intelligence in Education, eds. by H.C. Lane, K. Yacef, J. Mostow, and P. Pavlik, pp.451–462, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [11] 石井隆稔, 植野真臣他, “最大クリーク問題を用いた複数等質テスト自動構成手法,” 信学論 (D), vol.J97-D, no.2, pp.270–280, Feb. 2014.
- [12] T. Ishii, P. Songmuang, and M. Ueno, “Maximum clique algorithm and its approximation for uniform test form assembly,” IEEE Trans. Learning Technologies, vol.7, no.1, pp.83–95, 2014.
- [13] E. Tomita, S. Matsuzaki, A. Nagao, H. Ito, and M. Wakatsuki, “A much faster algorithm for finding a maximum clique with computational experiments,” J. Information Processing, vol.25, pp.667–677, 2017.
- [14] C.M. Li, H. Jiang, and F. Many’a, “On minimization of the number of branches in branch-and-bound algorithms for the maximum clique problem,” Comput. Oper. Res., vol.84, pp.1–15, 2017.
- [15] T. Ishii and M. Ueno, “Algorithm for uniform test assembly using a maximum clique problem and integer programming,” Int. Conf. Artificial Intelligence in EducationSpringer, pp.102–112, 2017.
- [16] 石井隆稔, 赤倉貴子, 植野真臣, “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,” 信学論 (D), vol.J100-D, no.1, pp.47–59, Jan. 2017.
- [17] 淵本 亮真, 植野真臣, “等質テスト構成における整数計画法を用いた最大クリーク探索の並列化,” 信学論 (D), vol.J103-D, no.12, pp.881–893, Dec. 2020.
- [18] H. Wainer, “Cats: Whither and whence,” Psicologica, vol.21, no.1, pp.121–133, 2000.
- [19] T. Ishii and M. Ueno, “Clique algorithm to minimize item exposure for uniform test forms assembly,” Int. Conf. Artificial Intelligence in EducationSpringer, pp.638–641 2015.
- [20] E. Boekkooi-Timminga, “The construction of parallel tests from irt-based item banks,” J. Educational Statistics, vol.15, no.2, pp.129–145, 1990.
- [21] R.D. Armstrong, D.H. Jones, and Z. Wang, “Automated parallel test construction using classical test theory,” J. Educational Statistics, vol.19, no.1, pp.73–90, 1994.
- [22] W.J. van der Linden and J.J. Adema, “Simultaneous assembly of multiple test forms,” J. Educational Measurement, vol.35, no.3, pp.185–198, 1998.
- [23] F.M. Lord and M.R. Novick, Statistical theories of mental test scores, IAP, 2008.
- [24] F.B. Baker and S.H. Kim, Item response theory: Parameter estimation techniques, CRC Press, 2004.
- [25] R.D. Armstrong, D.H. Jones, and C.S. Kuncze, “Irt test assembly using network-flow programming,” Applied Psychological Measurement, vol.22, no.3, pp.237–247, 1998.
- [26] IBM, “Ilog cplex optimization studio cplex user’s manual 12.9,” 2019.

(2021 年 7 月 27 日受付, 2022 年 1 月 24 日再受付, 4 月 19 日早期公開)



植野 晶

2021 電気通信大学情報理工学域卒。同年、電気通信大学大学院情報理工学研究科情報・ネットワーク工学専攻入学、現在に至る。e テスティング、等質テストの研究・開発に従事。



淵本 亮真

2020 電気通信大学情報理工学域・先端工学基礎課程卒。同年、電気通信大学大学院情報理工学研究科情報・ネットワーク工学専攻入学、現在に至る。e テスティング、等質テストの研究・開発に従事。



植野 真臣 (正員)

1992 神戸大学大学院教育学研究科了, 1994 東京工業大学大学院総合理工学研究科了。博士 (工学), 東京工業大学, 千葉大学, 長岡技術科学大学を経て 2006 より電気通信大学勤務, 2013 教授。2016 電気通信大学大学院情報理工学研究科教授, 現在に至る。電気通信大学大学院情報システム学研究科教授。人工知能, e テスティング, e ラーニング, ベイズ統計の研究に従事。2008 The 20th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2008) Best Paper Award. 2017 電子情報通信学会論文賞。