

クリーク列挙による項目露出を考慮した等質テスト構成

2022 年 2 月 28 日

情報数理工学プログラム

学籍番号 1810458

中島 太輔

指導教員 植野 真臣

令和3年度 情報数理工学プログラム卒業論文概要

平成30年度入学	学籍番号 1810458
指導教員 植野 真臣	氏名 中島 太輔
題目	クリーク列挙による項目露出を考慮した 等質テスト構成

概要

等質テスト構成のためにクリークを列挙し、項目露出が抑えられたクリークを選択する手法を提案する。等質テストとは、各テストに含まれるテスト項目は異なるが、受検者の推定得点の予測誤差が等質なテスト群である。項目がテスト構成全体で出題される回数のことを露出数と呼ぶが、H. Wainer(2000)によると露出数が多い項目は受験者間で共有されやすく、その項目の信頼性低下につながる事が知られている。石井ら(2017)は等質テスト構成を最大クリーク問題と整数計画法により、従来手法より多くのテストを構成できる手法を提案した。また、植野ら(2022)は最大クリークのテスト生成の際に整数計画法を適用することで、より露出数の小さくなる等質テスト構成の手法を提案した。しかし、クリークのサイズが等しい場合、初めに探索されたクリークが選択されている。

提案手法ではクリークを列挙するために生成されたテストを露出数の総和の順にテストを1つ選択する。そのテストに隣接したテストからなる部分グラフを抽出し、最大クリーク探索を行うことで等質テストを複数を構成し、構成数と最大露出数の差が最大の等質テストを出力する。

本論文ではシミュレーション及び実データを用い、テストをランダムと植野らの手法の2通りで生成し、提案手法の有効性を示した。

目次

1	まえがき	4
2	項目反応理論	5
3	先行研究	6
3.1	等質テストのための最大クリーク問題	6
3.2	乱択法	8
3.3	整数計画法を用いた等質テスト構成	9
3.4	露出数を考慮したテスト構成	10
4	提案手法	11
5	実験	15
5.1	RndMCP 法, 植野らの手法 [17] との比較実験	15
5.2	植野らの手法 [17] を用いて構成したテストと選択された等質テストの 関係	18
5.3	ランダムに構成したテストと選択された等質テストの関係	19
5.4	$OC = 5$ での等質テストの列挙数 U_v の最適化	22
6	むすび	22

1 まえがき

e テスティングとは、異なる問題で構成されるが、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, 2]。e テスティングでは一般的に”等質テスト”が用いられている。”等質テスト”とは出題項目は異なるが、項目数や同一能力者の得点予測の誤差がテスト間で等しいテスト群のことである。e テスティングの普及により、等質テストを自動生成する手法が多く提案されている [3, 4, 5, 6]。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される項目や統計データが格納されており、これらの項目から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生成させることによってテストを構成する。図 1 が等質テストの自動構成の概念図となる。

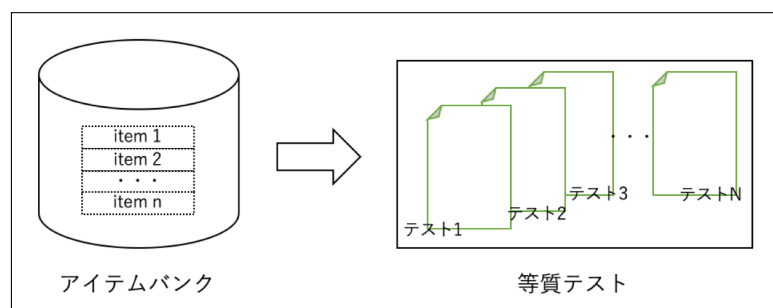


図 1: テストの自動構成の概念図

テストの自動構成は数理最適化問題として解かれる。先行研究として 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。この手法は情報処理技術者試験をはじめとした、我が国の国家試験において実際に使用されている [5]。石井ら (2014) はグラフからランダムに部分グラフを抽出し、最大クリーク探索をすることでグラフ全体の最大クリークを近似的に探索する等質テスト構成手法 (RndMCP 法) を提案した [7]。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できた [8]。しかし最大クリーク探索はグラフの頂点集合を V とすると、最先端のクリーク探索手法 [9, 10] 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るという問題があった。

RndMCP 法の空間計算量を緩和するために、石井ら (2017) は第一段階で RndMCP 法を用いてメモリ限界までの最大クリーク探索を行い、第二段階で第一段階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必要な計算量を $O(|V|)$ に削減する HybridRBP 法を提案した [11, 12]。これにより 10 万をこえる等質テストを構成することが可能となった。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僅かであった。そのため、渕本, 植野 (2020) は HybridRBP 法 [11, 12] を並列探索アルゴ

リズムに改善し、さらに多くのテスト構成を可能とした [13]. 先行研究 [8, 11, 12] ではある項目がテスト構成全体で出題される回数 (以下, 露出数と呼ぶ) に偏りが生じる問題がある. 露出数の大きい項目は多くのテストで出題されているため, 受験者間で共有されやすく, 項目の信頼性の低下につながる [14]. この露出数の偏りを軽減するために石井ら (2015) は RndMCP 法と整数計画法を用いてテストを生成し, その中から項目露出率 (= 露出数の最大値/生成されたテストの数) が最も小さいものを選択する手法を提案した [15]. これにより, 従来手法よりも露出率を軽減することができた.

従来手法 [16] の第一段階目の 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テスト構成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, 植野らは整数計画法を用いてテスト構成する手法を提案した [17]. 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テストを構成することができた. これにより, 第一段階目のテスト構成において項目露出の偏りを抑制した. 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解決されない. そこで露出数上位 s 位までの項目を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した. 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, 最大露出数が抑制できた. 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ため, その際は除外した項目をすべてアイテムバンクに戻す. この手法によって, 露出率だけでなく, テスト構成数も改善できた.

しかし, 等質テストの構成数が同一の場合であっても, 最大露出数が異なる場合がある. また, 植野ら [17] の手法では, 項目露出をテスト構成の際に考慮するが, 最大クリーク探索時には考慮しない. したがって, 露出数を考慮する等質テスト構成では最適ではない. この問題を解決するために本論文では, 露出数の総和の順にテストを並べ替え, その順にそのテストを必ず含む等質テストを構成し, 繰り返し行うことで等質テストを複数構成する. その中から最大露出数が小さく, テスト構成数が大きくなる等質テストを選択する手法を提案する.

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. 提案手法では, 従来手法と同程度のテスト構成数を保ちながら, 露出率を改善できた. また, クリーク探索時間が同時間では一部の条件においてテスト構成数が改善された. 提案手法や先行研究 [8, 17] では, 最大クリーク探索に制限時間を設けているが, 項目露出数でテストを並び替えてから探索することで, 従来手法よりも大きなクリークを早期に発見できた.

2 項目反応理論

一般的に等質テストは以下の条件を満たすテストの集合として定義される [6, 7, 8].

1. それぞれのテストで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数以下である．(以降、項目重複数条件と呼ぶ)

受験者の予測得点の誤差はテストの自動構成に関する研究 [3, 5, 23, 18, 19] では項目反応理論 (Item Response Theory:IRT) [20, 21] におけるテスト情報量を用いて評価されている．IRT とは受験者の各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、異なる項目で構成されたテストであっても受験者能力の同一尺度での比較が可能となる．

テスト情報量を扱う際に用いられる IRT は 2-パラメータロジスティックモデル (2-Parameter Logistic Model:2PLM) であり、受験者 $j = (1, \dots, m)$ が項目 $i = (1, \dots, n)$ に正答する確率 $p_i(\theta_j)$ は以下のようにモデル化される．

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

ここで $\theta_j \in (-\infty, \infty)$ は受験者 j の能力パラメータ、 $a_i \in [0, \infty]$, $b_i \in [0, \infty]$ はそれぞれ i 番目の項目の識別パラメータ、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情報量の総和をテスト情報量と呼び、以下のように定義される．

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

ここでの 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, 2]．

ただし、テストの自動構成手法 [3, 5, 23, 18, 19, 22] ではテスト情報量における受験者の能力パラメータ θ_i を $\Theta = \{\theta_1, \dots, \theta_k, \dots, \theta_K\}$ のように離散的にサンプリングしている．

3 先行研究

本章では提案手法と関連がある手法を紹介する．

3.1 等質テストのための最大クリーク問題

石井ら (2014) は等質テストの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 [7]．ここでクリークとはグラフ上の任意の 2 頂点間が隣接する頂点集合である．

表 1: テスト情報量への上下限制約の例

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

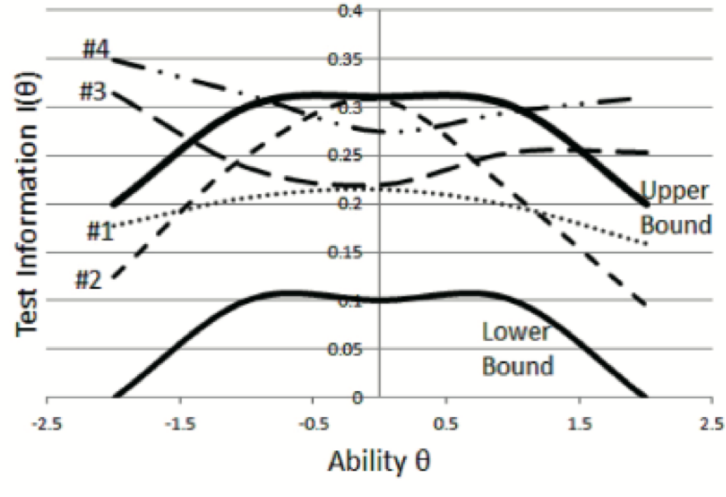


図 2: テスト情報量への上下限制約の例

本手法では能力パラメータ $\theta = \{\theta_1, \dots, \theta_k, \dots, \theta_K\}$ を各点ごとにテスト情報量 $I(\theta_k)$ の上下限制約 ($UB(\theta_k), LB(\theta_k)$) を設定し、全ての上下限制約を満たすテストは受験者の得点予測誤差が等質であるとする。例えば、図 2 は表 1 のテスト情報量の制約を与えた時の概念図である。図 2 中、#1, #2 のテストは各点で上下限制約を満たすので等質なテストである。一方で #3, #4 は上限制約を満たしていない箇所が存在するので等質テストではない。

生成されるテスト候補を以下のグラフ構造とみなし、クリークを探索する。

1. 頂点：項目重複数条件以外を満たす全てのテストを頂点とする。
2. 辺：2 頂点に対応するテストが項目重複条件を満たす場合に辺をひく。

グラフ上の任意の頂点はテスト構成条件を満たし、クリーク中の 2 頂点は隣接しており、項目重複条件を満たす。したがって、クリーク中に含まれる頂点であるテストはそれぞれ等質テストであり、頂点数が最大となるクリークが最大の等質テストである (図 3)。

等質テストの構成は頂点の集合を V 、辺の集合を E とする無向グラフ $G = (V, E)$ 、クリークを C とすると以下のように定式化できる。

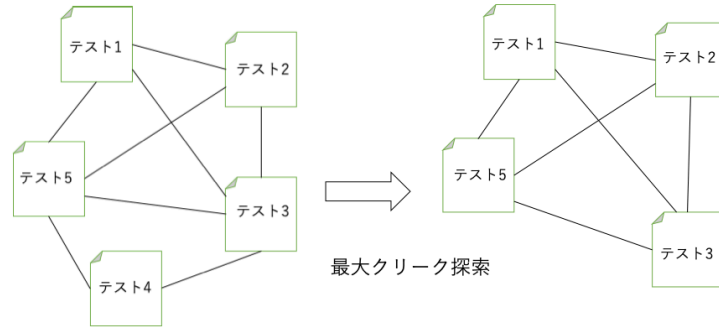


図 3: 等質テストの略図

複数等質テストの最大クリーク問題

variables $C \subseteq V$

maximize $|C|$

subject to

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

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

この手法により理論上、等質テストを最大個数構築することが可能だが、実行に計算時間量 $O(2^{|V|})$ 、空間計算量 $O(|V|^2)$ が必要である。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、これはアイテムバンクに含まれる項目数 n と連動して n が増加するにつれて爆発的に増えていく。そのために現在用いられている数百から千以上の項目を持つアイテムバンクでは等質テストの構成を厳密に行うことは困難である。

3.2 乱択法

これらの計算コストの問題を緩和するために石井ら（2014）は擬似的な最大クリーク探索を行うアルゴリズム（RndMCP 法）を提案した [8]。RndMCP 法ではテスト構成数に上限を与えることで部分グラフをランダムに構築し、最大クリーク探索を繰り返し行うことにより、グラフ全体の最大クリークを近似的に探索を行う。具体的なアルゴリズムは Algorithm1 に示す。Step1 において項目重複数以外の条件を満たすテストを L_1 個構成し、Step2 にて項目重複数条件を満たす 2 頂点間を辺で結ぶことで部分グラフを構築する。ただし、 L_1 はチューニングパラメータであり、用いる計算機のメモリの上限に応じて設定する。Step3 では構築した部分グラフの最大クリーク探索を L_2 時間だけ行う。Step1~3 の合計時間を CT 時間とし、この時間内で Step1~3 を繰り返し行うことで最大クリークを出力する。

本手法は最大クリーク探索の時間・空間計算量が $O(2^{|V|})$, $O(|V|^2)$ となった。この時、探索するグラフの頂点数 $|V|$ は L_1 、探索時間は L_2 としてパラメータにすることができた。従って、現実時間で計算機のメモリ上限まで最大クリーク探索を行えるようになった。これにより一般的な規模のアイテムバンクから最大 10 万頂点ほど

の等質テストを生成できた.

Algorithm 1 乱択法

Require: アイテムバンク, テスト構成条件

Ensure: 等質テスト群

```
1: PROCEDURE RndMCP( $L_1, L_2, CT$ )
2:  $C = \emptyset, C_{\max} = \emptyset$ 
3:  $ST = currenttime$ 
4: while ( $currenttime - ST$ ) <  $CT$  do
5:   /*Step1*/
6:    $V = L_1$ 個のテストをランダム生成▷項目重複数条件以外を満たす  $L_1$ 個のテスト
7:   /*Step2*/
8:    $G = (V, E)$ : グラフ構築▷二頂点が重複項目数条件を満たす場合, 辺を引く
9:   /*Step3*/
10:   $C = MCP(G, L_2)$ ▷ $L_2$ 時間で最大クリーク探索を行う
11:  if  $|C_{\max}| \leq |C|$  then
12:     $C_{\max} = C$ 
13:  end if
14: end while
15: return  $C_{\max}$ 
16: ENDPROCEDURE
```

3.3 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 [9, 10] を用いたとしても空間計算量を $O(|V|^2)$ だけ必要とするため 10 万個程度のテスト構成が上限であった. そこで石井ら (2017) は RndMCP 法と整数計画法を組み合わせた二段階探索手法 (HybridRBP 法) を提案した [11, 12].

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以下の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, この探索法は $O(|V| \cdot 2^n)$ だけ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よりも大幅に劣る. そこで RndMCP 法により計算機のメモリ限界の頂点数 L_1 を持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 [11, 12]. 以下の整数計画法でクリークに隣接する頂点を探索する.

制約条件はクリーク 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

るたび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 and Armstrong (2006) 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 [6].

最大クリーク探索のための整数計画法

variables

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

maximize

$$\sum_{i=1}^n \lambda_i x_i \quad (4)$$

subject to

$$\sum_{i=1}^n x_i = M(\text{テスト項目数}) \quad (5)$$

$$LB_{\theta_k} \leq \sum_{i=1}^n I_i(\theta_k) x_i \leq UB_{\theta_k} \quad (6)$$

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

$$\sum_{i=1}^n X_{i,r} x_i \leq OC(\text{項目重複上限数}) \quad (7)$$

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

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

3.4 露出数を考慮したテスト構成

RndMCP 法のテスト構成における露出数の偏りが乗じる問題がある．そこで植野ら (2021) は整数計画法を用いてテスト構成する手法を提案した [17]. 本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 (4)) 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 は一様分布に従う乱数であるため、制約を満たすテスト全体からテストが一様に構成される．このため、第一段階目のテスト構成において項目露出の偏りを改善することができた．さらに、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、テスト構成を行った．

具体的には、テスト構成において項目重複数条件を除いた整数計画問題 (式 (4) ~ (7)) を解き、テスト構成し追加する．テスト構成後に、全ての項目のうち構成したテスト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．以上をテスト構成数が L_1 になるまで逐次的に行い、テストを構成する．ただし、新た

にテストを構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.
この手法のアルゴリズムを Algorithm2 に示した.

Algorithm 2 露出数を考慮したテスト構成

Require: アイテムバンク, テスト構成条件

Ensure: テスト群

```
1: PROCEDURE LimitItemRndMCP( $L_1, CT, s$ )
2:  $ST = currenttime$ 
3: while ( $currenttime - ST$ ) <  $CT$  do
4:    $V = \emptyset$ 
5:    $n = itemBank.size$ 
6:   for  $i \leftarrow 0$  to  $n - 1$  do
7:      $IEV[i] = 0$ 
8:   end for
9:    $allowedItems = \text{アイテムバンクの全項目}$ 
10:  while  $|V| \leq L_1$  do
11:     $Sol = \text{IPsolve}(allowedItems)$  /* $allowedItems$  のみからなるアイテムバンク
    で式 (4)~式 (6) の IP を解く*/
12:    if  $Sol = \emptyset$  then
13:       $allowedItems = \text{アイテムバンクの全項目}$ 
14:      break
15:    end if
16:     $V = V \cup \{Sol\}$ 
17:    for  $i \in Sol$  do
18:       $IEV[i]++$ 
19:    end for
20:     $allowedItems = \text{IEV の上位 } s \text{ 位以外の項目}$ 
21:  end while
22: end while
23: return  $V$ 
24: ENDPROCEDURE
```

4 提案手法

先行研究 [7, 8, 11, 12, 14, 15] では第一段階目の RndMCP 法で構成されたテストの露出数の偏りが大きく、石井ら (2015) の手法を用いても露出数の改善に限界があった. そこで、植野ら (2021) はテスト構成する際に露出数をペナルティとして扱うことで大幅に改善することができた [17]. しかし、第一段階目に行われる最大クリーク探索で得られた等質テスト構成数が同一の場合であっても、最大露出数が

異なる場合がある．従来手法では先に探索された最大クリークを構成数最大の等質テストとして扱い，同一の構成数の等質テスト間で露出数が考慮されていない．この問題を解決するために，本研究ではテスト構成数最大の等質テストを列挙し，その中から等質テスト構成数と最大露出数の差が最も大きいものを選択する手法を提案する．しかし，初めに構成したテスト数に応じて等質テストの列挙には時間計算量，空間計算量が十分量必要となる．この問題を解決するために，テストを露出数の総和の昇順に並べ替え，指標の値の順にそのテストが必ず含む等質テストを構成していく．これにより，列挙回数を所定の回数で打ち切って露出数とテスト構成数を最適化できる．

具体的には，構成されたテストに対して，指標の昇順に並べ替えることでより露出率が小さくなるテストを判別する $(t_1, t_2, \dots, t_{|V|})$ ．指標が最も小さいテストの順に，そのテスト t_s に隣接したテストからなる部分グラフを抽出して最大クリーク探索を行う． s の値を順に U_v まで変更していき，それぞれの隣接グラフにおける最大クリークを列挙していくことで所定の時間内でクリークの列挙を行う (U_v はチューニングパラメータ)．

提案手法は概要図 4 のように，以下の手順に従いクリークを決定する．

- (1) テストを $|V|$ 個構成 (植野 2021 の手法を利用)
- (2) テストを露出数の総和の昇順に並べ替える $(t_1, t_2, \dots, t_{|V|})$
- (3) テストを頂点，等質条件を満たす 2 頂点間に辺を結び，グラフを生成
- (4) テスト t_s に隣接したテストからなる部分グラフを抽出
- (5) 部分グラフに対して最大クリーク探索を行う
- (6) 得られた等質テスト (最大クリークに属するテスト + t_s) の構成数と最大露出数の差を計算
- (7) 6 の値が最大の等質テストを保存

並べ替えの指標は，生成されたグラフの頂点集合 (テスト集合) を V とし，項目 i の V における露出数 IEV_i ，頂点 t における最大露出数 $IER_{\max t}$ ，最小露出数 $IER_{\min t}$ ，露出数の総和 $IER_{\text{sum} t}$ であり，それぞれ以下のように定義する．

$$IEV_i = \sum_{t=1}^{|V|} Y_{i,t} \quad (8)$$

$$Y_{i,t} = \begin{cases} 1 & (t \text{ 番目のテストに項目 } i \text{ が含まれる}) \\ 0 & (t \text{ 番目のテストに項目 } i \text{ が含まれない}) \end{cases} \quad (t = 1, \dots, |V|)$$

$$IER_{\max t} = \max_{i \in t} IEV_i \quad (9)$$

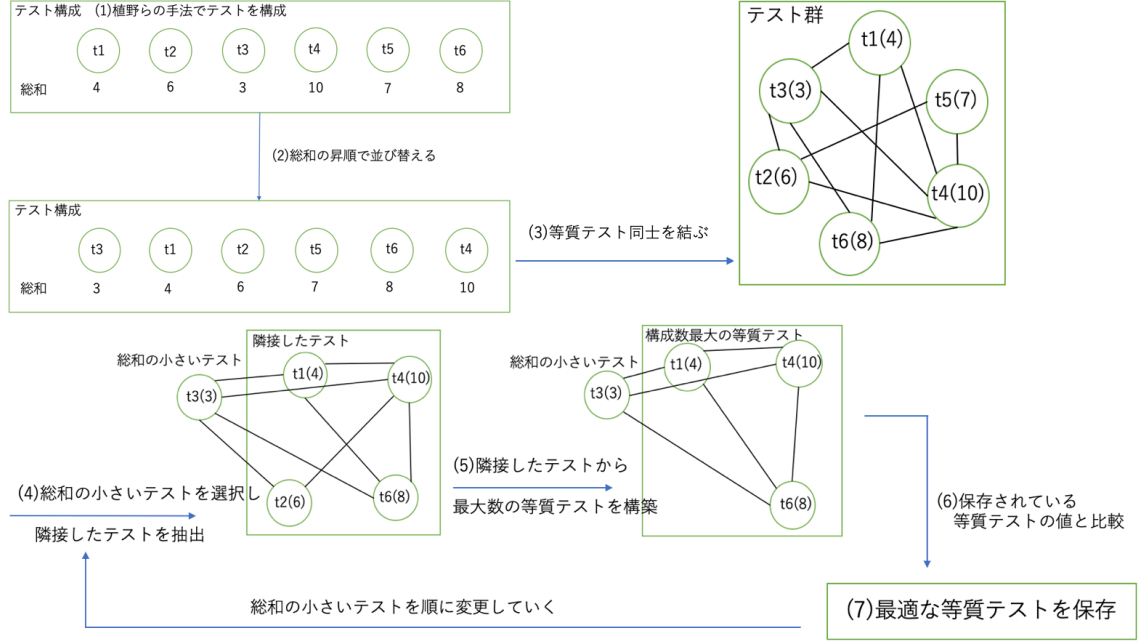


図 4: 提案手法の概念図

$$IERmin_t = \min_{i \in t} IEV_i \quad (10)$$

$$IERsum_t = \sum_{i=1}^n IEV_i \times Y_{i,t} \quad (11)$$

で表す。テスト構成 C における項目 i の露出数 IEC_i は以下のように表される。

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

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

また、テスト構成 C における最大露出数を

$$IEC_{\max} = \max_{i=1, \dots, n} (IEC_i) \quad (13)$$

で表す。このときテスト構成 C における項目露出率は

$$IEC_{rate} = \frac{IEC_{\max}}{|C|} \quad (14)$$

クリークサイズ、項目露出数の両方を考慮したものは

$$IEC_{both} = |C| - IEC_{\max} \quad (15)$$

の値が大きいものを選択している。

本手法により、露出数の小さい項目からなるテストでの複数等質テストを生成す

ることができる．提案手法のアルゴリズムを Algorithm3 に示す．

step1 では植野らの手法を用いてテストを L_1 個構成する．

step2 では step1 で構成した L_1 個のテストを最大露出数，最小露出数，露出数の総和の昇順に並べ替える．その頂点集合を V とし，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 では指標の昇順にテスト t_s を t_{U_v} まで逐次的に変更していき， t_s に隣接するテストからなる部分グラフの最大クリーク探索を行う．その際，最大クリーク探索の時間を L_2 時間とする． s を変更するときに選択基準の値を計算し， C_{best} の変更を同時に行い，等質テスト構成をする．

Algorithm 3 クリーク列挙アルゴリズム

Require: アイテムバンク, テスト構成条件

Ensure: 等質テスト群

```
PROCEDURE MCP( $L_1, L_2, U_v$ )
   $C = \emptyset, C_{best} = \emptyset$ 
3: /*Step1*/
   $V = L_1$  個のテストを整数計画法を用いて生成
  /*step2*/
6: for  $x \leftarrow 0$  to  $L_1 - 1$  do
  /*各指標の値を計算する (例:最大値)*/
   $max[x][0] \leftarrow x$  番目に生成されたテスト
9:    $max[x][1] \leftarrow IERmax_x$ 
  end for
  Sorting var $t_x$ 
12: for  $x \leftarrow 0$  to  $L_1 - 1$  do
   $V \leftarrow max[x][0]$ 
  end for
15:  $G = (V, E) \triangleright 2$  頂点が重複項目数条件を満たす場合，辺を引く
  /*Step3*/
  while  $s \leftarrow 0$  to  $U_v$  do
18:   /*Step3.1*/
   $g = (v, e, t_s) \triangleright g$  は  $t_s$  に隣接する頂点からなる部分グラフ;  $v \subseteq V, e \subseteq E$ 
   $C = MCP(g, L_2) \triangleright g$  の最大クリークを  $L_2$  時間だけ探索
21:   /*Step3.2*/
  if  $|C_{best}| - E_{C_{best}} < |C| - E_C$  then
     $C_{best} = C$ 
24:   end if
  end while
  return  $C_{best}$ 
27: ENDPROCEDURE
```

表 2: 実アイテムバンクの詳細

Pool Size	Parameter a			Parameter b		
	Range	Mean	SD	Range	Mean	SD
978	0.12 ~ 3.08	0.43	0.2	-4 ~ 4.55	-0.22	1.16

表 3: 本実験のテスト情報量の上下限制約

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

5 実験

提案手法の有効性を示すために、それぞれのクリーク列挙から選択された等質テストと従来手法 RndMCP 法との比較実験を行った。実行環境は Ubuntu20.04.3LTS を OS とする計算機（CPU：Intel Core i9-10900k 3.70GHz，RAM：64GB）である。従来手法の結果は植野らの手法の $s = 1$ であり、これは露出数の上位 1 位の項目を除いたものである。グラフを構築する際の頂点集合 V は同じものを使用している。本実験では実データとシミュレーションデータによるアイテムバンクを用いた。実アイテムバンクは 978 項目をもち、そのパラメータの詳細は表 2 の通りである。シミュレーションアイテムバンクは 500, 1000, 2000 項目を持ち、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 $b \sim N(0, 1^2)$ として発生させた。

テストの構成条件は前述したアイテムバンクから、表 3 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 [17, 16] と同様に $OC = \{0, 5, 10\}$ の 3 通りの条件で評価する。

また、比較項目として以下の項目を用いた。

各項目の露出数 IEC_i の平均を

$$IEC_\mu = \frac{\sum_{i=1}^n IEC_i}{n} \quad (16)$$

とし、各項目の露出数 IEC_i の標準偏差を

$$IEC_\sigma = \sqrt{\frac{\sum_{i=1}^n (IEC_i - IEC_\mu)^2}{n}} \quad (17)$$

項目重複数を OC ，テスト構成数を $|C|$ ，最大露出数を IEC_{max} ，露出率を IEC_{rate} ，露出数の標準偏差 IEC_σ ，等質テストの選択基準を IEC_{both} とした。

5.1 RndMCP 法，植野らの手法 [17] との比較実験

HybridRBP 法の第一段階目に構成されるに等質テストにおいて、等質テスト構成数と露出数のトレードオフの関係を改善できることを示すために、RndMCP 法，整数計画法のみと提案手法との比較実験を行った。本比較実験ではテストの並べ替えの

表 4: 実験結果

Pool Size	OC	従来手法 (RndMCP)					従来手法 (整数計画法) [17]				
		$ C $	IEC_{σ}	IEC_{max}	IEC_{rate}	IEC_{both}	$ C $	IEC_{σ}	IEC_{max}	IEC_{rate}	IEC_{both}
500	0	10	0.5	1	0.100	9	11	0.5	1	0.0909	10
	5	4371	53.1	378	0.0865	3993	4994	50.2	379	0.0759	4615
	10	99981	1729.2	13006	0.130	86975	99970	1138.4	5370	0.0537	94600
1000	0	17	0.5	1	0.0588	16	18	0.5	1	0.0556	17
	5	46190	430.5	3374	0.0730	42816	50779	377.7	1616	0.0318	49163
	10	100000	990.7	8767	0.0877	91233	99998	769.5	2847	0.0285	97151
2000	0	32	0.5	1	0.0313	31	32	0.5	1	0.0313	31
	5	96773	456.1	3833	0.0396	92940	97320	377.9	1397	0.0144	95923
	10	100000	472.6	4044	0.0404	95956	100000	388.8	1418	0.0142	98582
978(actual)	0	18	0.5	1	0.0556	17	19	0.5	1	0.0526	18
	5	45790	368.2	5177	0.113	40613	54800	204.6	1683	0.0307	53117
	10	100000	985.3	16310	0.163	83690	100000	418.7	2754	0.0275	97246

Pool Size	OC	提案手法									
		max					min				
		$ C $	IEC_{σ}	IEC_{max}	IEC_{rate}	IEC_{both}	$ C $	IEC_{σ}	IEC_{max}	IEC_{rate}	IEC_{both}
500	0	11	0.5	1	0.0909	10	11	0.5	1	0.0909	10
	5	4962	49.9	374	0.0754	4588	4985	50.2	386	0.0774	4599
	10	99971	1139.6	5370	0.0537	94601	99971	1139.6	5370	0.0537	94601
1000	0	18	0.5	1	0.0556	17	18	0.5	1	0.0556	17
	5	50759	377.7	1641	0.0323	49118	50790	374.7	1661	0.0327	49129
	10	99999	769.8	2847	0.0285	97152	99999	769.8	2847	0.0285	97152
2000	0	32	0.5	1	0.0313	31	32	0.5	1	0.0313	31
	5	97321	378.0	1398	0.0144	95923	97321	377.0	1399	0.0144	95922
	10	100000	388.8	1418	0.0142	98582	100000	388.8	1418	0.0142	98582
978(actual)	0	19	0.5	1	0.0526	18	19	0.5	1	0.0526	18
	5	54881	205.5	1678	0.0306	53203	54861	197.9	1737	0.0317	53124
	10	100000	419.0	2754	0.0275	97246	100000	419.0	2754	0.0275	97246

指標として最大露出数の昇順 max と最小露出数の昇順 min を用いた。パラメータ値は $L_1 = 100000$, $L_2 = 2hr$, $U_v = 2$ とした時の結果を表 4 に示す。

表 4 の結果より, $OC = 0$ の時は全てのアイテムバンクにおいて提案手法を用いても整数計画法と同じ大きさの等質テストが得られた。このことから, $OC = 0$ の場合において 2 時間のクリーク探索であっても同じ大きさの等質テストを得られることがわかった。そのため, 少なくとも 2 時間で HybridRBP 法における第一段階を終えることができると本実験よりわかった。グラフが疎であればあるほど部分グラフの大きさが小さくなる。以下の表 5 は各アイテムバンクにおける部分グラフの大きさである。これより提案手法を用いることで従来手法よりも探索すべきグラフを小さくできるため半分の時間で同じ大きさの等質テストを得られることができた。 $OC = 10$ において, 構成されたテストからなるグラフの密度は以下の表 6 となる。

表 5: 各アイテムバンクにおける $OC = 0$ の探索する部分グラフの大きさ

Pool Size	従来手法 (整数計画法) [17]	提案手法			
		max		min	
		1 回目	2 回目	1 回目	2 回目
500	100000	28051	27824	28086	27150
1000	100000	54189	54271	51569	51113
2000	100000	73066	73155	72108	71487
978	100000	57599	56607	57047	56109

表 6: 各アイテムバンクにおける $OC = 10$ のグラフ密度

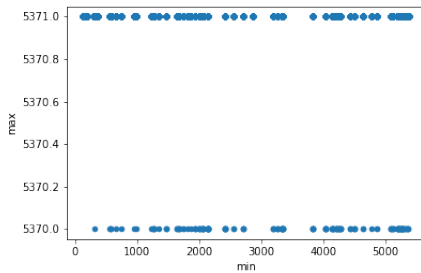
Pool Size	グラフの辺数	G_D
500	4999949968	0.9999999936
1000	4999949999	0.9999999998
2000	4999950000	1.000000000
978	4999950000	1.000000000

グラフの密度 G_D は以下のように求めた.

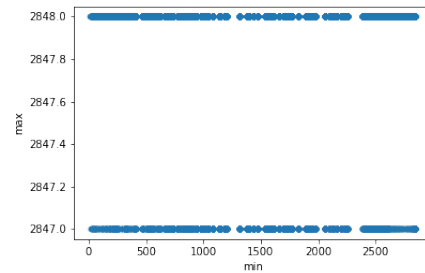
$$G_D = \frac{\text{実際のグラフの } edge \text{ 数}}{|V|C_2}$$

これにより元々のグラフが完全グラフに近く、部分グラフを抽出せずとも等質テストに含まれない無駄な探索が行われにくいため、提案手法との差異があまり見られなかった.

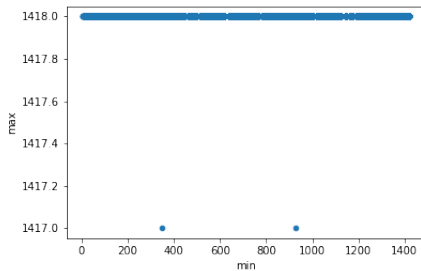
一方で、 $OC = 5$ の時ではシミュレーションデータにおける提案手法は有効な手法ではなかったが実データにおいては最も有効な手法であった. そこで各アイテムバンクにおける整数計画法で構成された各テストに含まれる項目の最大露出数と最小露出数の分布は以下の図5のようになった.



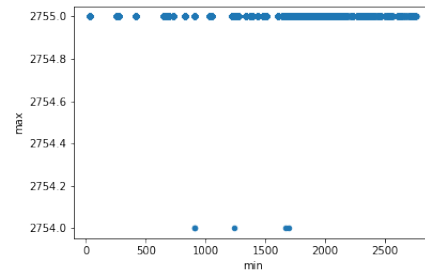
(a) $n=500$



(b) $n=1000$



(c) $n=2000$



(d) $n=978$

図 5: 各アイテムバンクにおける整数計画法を用いて生成された頂点の最大露出数と最小露出数の分布

図5より各テストに含まれる項目の最小露出数は一様に分布しているが、最大露出数は全てのテストが二値になっていることがわかる. 構成された各テストの最大

露出数にばらつきが小さい場合、最大露出数の昇順で並び替えることは効果的でないと思われる．そのため、今後の比較は $OC = 5$ かつ露出数の総和を指標としたものを分析する．

5.2 植野らの手法 [17] を用いて構成したテストと選択された等質テストの関係

提案手法におけるテスト t_i を露出数の総和の最小かつ同値の場合は最小値が小さいもの ($i=1,2$) とした場合、選択された等質テストは以下の表 7 のようになった．表 7 では $n = 500$ 以外では等質テストの構成数が大きくなった．さらに、項目の標準偏差は全てのアイテムバンクにおいて小さくなった．そこで等質テストに含まれるテストと構成されたテストの傾向を判断する．そのために、図 5 と同じテストによる各テストの露出数の総和と最小露出数の分布と提案手法によって得られた等質テストに含まれるテストの構成された時点での露出の総和と最小露出数の分布を図 6 に示す．

図 6 より、各アイテムバンクに対して最小露出数と露出数の総和がともに小さく

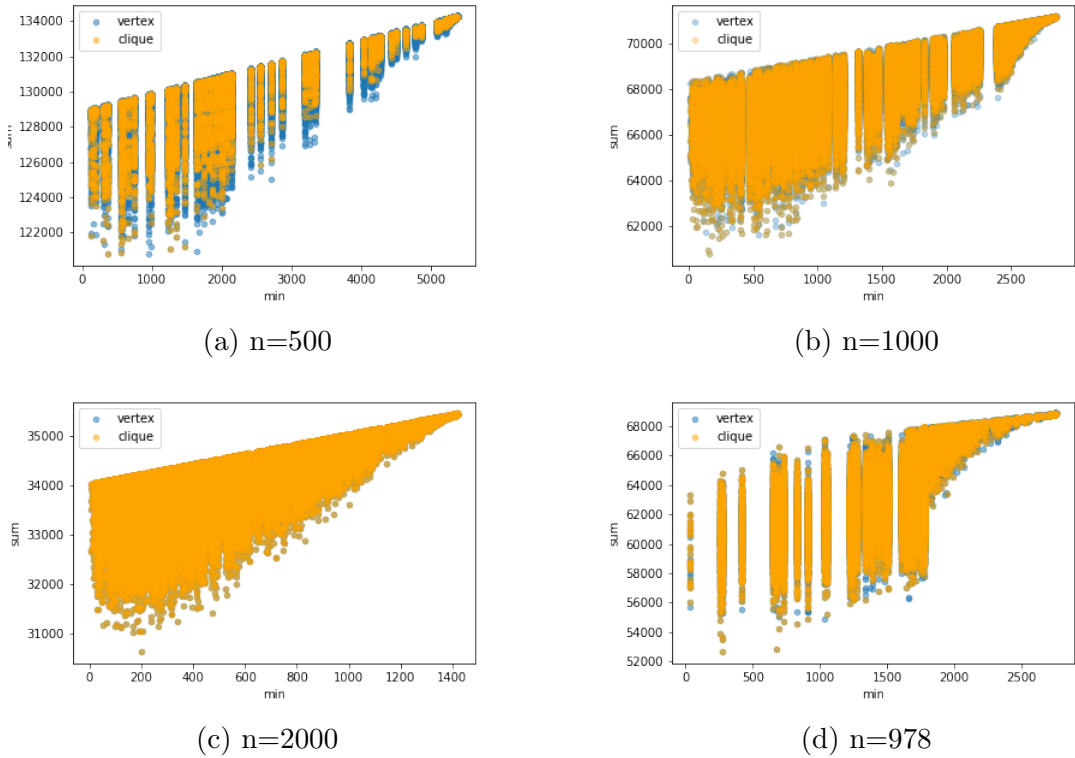


図 6: 各アイテムバンクにおける $OC = 5$ の構成される全テストと等質テストに含まれるテストの露出数の総和と最小露出数の分布

なるテストからそのテスト t_s に隣接するテストからなる部分グラフにおいてクリーク探索を行なった場合、 $n = 500, 1000$ においてテスト t_s と似た性質を持つテストが

表 7: $OC = 5$ における選択された等質テスト

Pool Size	従来手法（整数計画法）[17]					提案手法（露出数の総和）				
	$ C $	IEC_{σ}	IEC_{max}	IEC_{rate}	IEC_{both}	$ C $	IEC_{σ}	IEC_{max}	IEC_{rate}	IEC_{both}
500	4994	50.2	379	0.0759	4615	4972	49.9	370	0.0744	4602
1000	50779	377.7	1616	0.0318	49163	50792	374.2	1661	0.0327	49131
2000	97320	377.9	1397	0.0144	95923	97321	376.8	1399	0.0144	95922
978	54800	204.6	1683	0.0307	53117	54851	195.9	1732	0.0316	53119

表 8: 各アイテムバンクにおける閾値以下のテストからなるグラフ密度

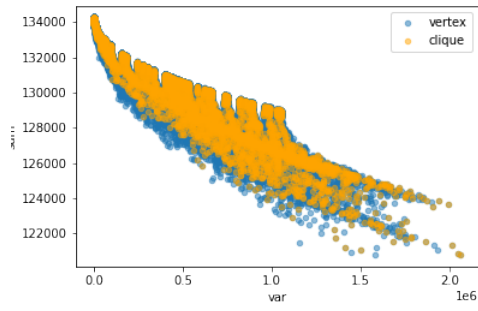
Pool Size	閾値	テスト数	グラフの edge 数	G_D
500	124000	319	50551	0.995
1000	64000	831	344837	0.999
2000	32000	352	61776	1.00
978	58000	65	2074	0.997

等質テストに含まれていないことがわかる．そこで，アイテムバンクごとに露出数の総和に閾値を設定し，その値以下のテストからなるグラフ密度を表 8 に示す．表 8 より，閾値以下のテストであっても密度が大きいとため，閾値外のテストと項目重複しているのでは構成数最大の等質テストに含まれていないと考えられる．また，露出数の総和が等しい場合，最小露出数が大きいテストほど等質テストに含まれにくいことが分かった．総和が等しく，最小値が大きいということは他の露出数との分散が小さいということである．そこで同じ等質テストに対して露出数の総和と各テストに含まれる項目の分散の分布を以下の図 7 に示す．これより，露出数の総和に対して分散が小さいテストほど等質テストに含まれていない．これは，テストに含まれる項目を構成の際に制限しているため，各項目の露出数の分散が小さい．多数の項目が多数のテストに含まれるが，項目数が小さいほど同じ項目が多くテストに含まれているために項目重複条件を満たしにくくなっていると考えられる．そのため，アイテムバンク数が大きくなるにつれてこの傾向が下がっていき， $n = 2000$ では完全グラフに近いグラフ構造になっている．

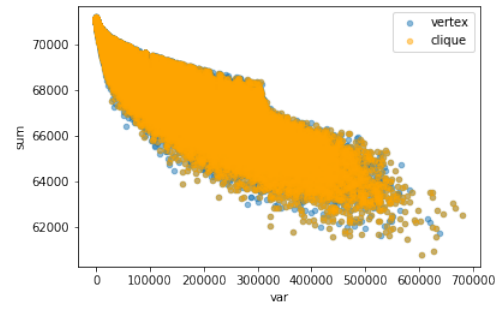
5.3 ランダムに構成したテストと選択された等質テストの関係

ランダムに構成したテストに対しても同様の実験を行なった．構成された全テストにおける露出数の最小値と最大値の関係は以下の図 8 のようになった．これらのテストに対して露出数の総和の最小かつ同値の場合は最小値が小さいものを各アイテムバンクに対して最小露出数と露出数の総和がともに小さくなるテストからそのテスト t_s に隣接するテストからなる部分グラフにおいてクリーク探索を行なった．その時の等質テストを表 9 に表し，図 9 に同じテストによる各テストの露出数の総和と最小露出数の分布と提案手法によって得られた等質テストに含まれるテストの構成された時点での露出数の総和と最小露出数の分布を示す．

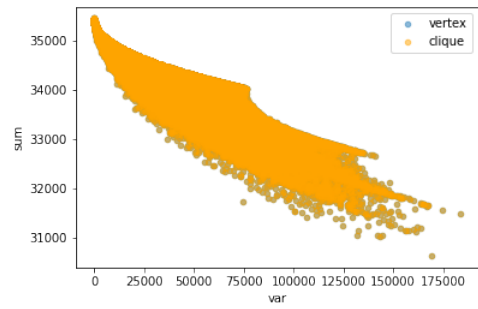
表 9 よりランダムに構成されたテストに対して，提案手法を用いることで等質テストの構成数以外の指標において改善することができた．これは図 9 の分布から構



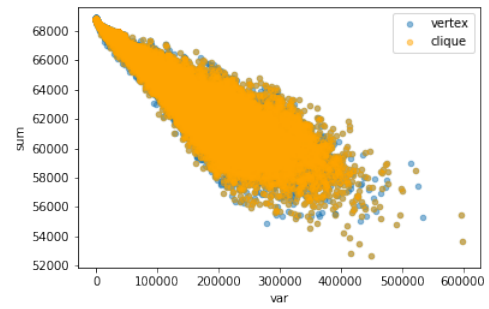
(a) $n=500$



(b) $n=1000$

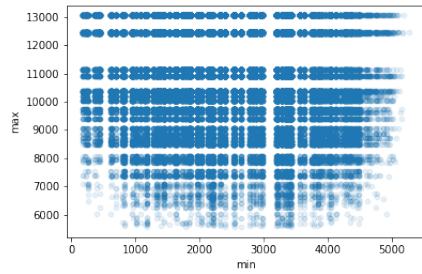


(c) $n=2000$

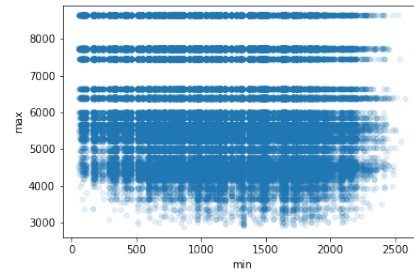


(d) $n=978$

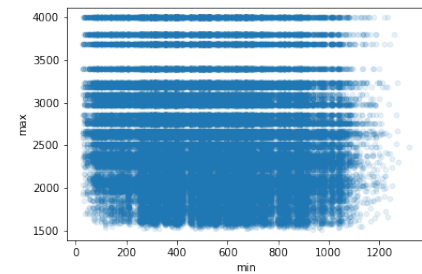
図 7: 各アイテムバンクにおける $OC = 5$ の構成されたテストと等質テストに含まれるテストの露出数の総和と分散の分布



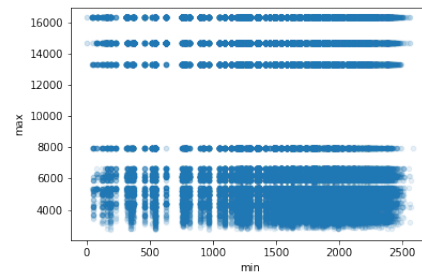
(a) $n=500$



(b) $n=1000$

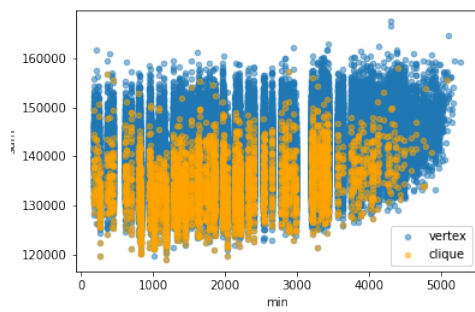


(c) $n=2000$

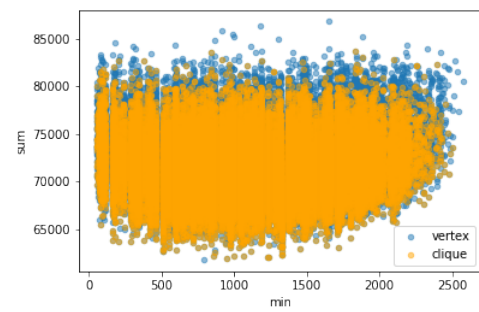


(d) $n=978$

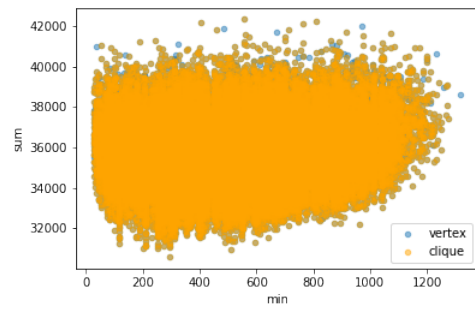
図 8: 各アイテムバンクにおけるランダムに構成された全テストの最大露出数と最小露出数の分布



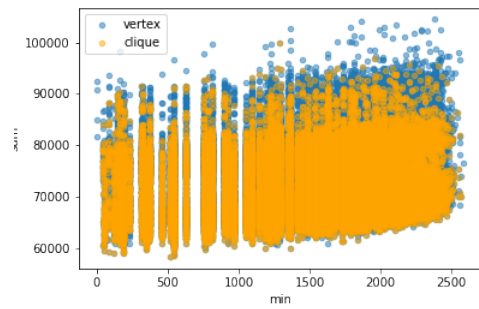
(a) $n=500$



(b) $n=1000$



(c) $n=2000$



(d) $n=978$

図 9: 各アイテムバンクにおけるランダムなテスト構成と等質テストに含まれるテスト露出数の総和と最小露出数の分布

表 9: $OC = 5$ におけるクリーク

Pool Size	従来手法 (rndMCP)					提案手法 (露出数の総和)				
	$ C $	IEC_{σ}	IEC_{max}	IEC_{rate}	IEC_{both}	$ C $	IEC_{σ}	IEC_{max}	IEC_{rate}	IEC_{both}
500	4373	53.4	382	0.0874	3991	4363	53.1	374	0.0857	3989
1000	46354	431.8	3326	0.0718	43019	46293	426.0	3215	0.0694	43078
2000	96623	455.2	3830	0.0396	92793	96623	453.0	3777	0.391	92846
978	45927	369.5	5110	0.111	40817	45898	362.0	4987	0.109	40911

表 10: 列挙数 U_v を変更した時の実験結果

Pool Size	U_v	提案手法 (露出数の総和)					従来手法 (整数計画法) [17]				
		$ C $	IEC_{σ}	IEC_{max}	IEC_{rate}	IEC_{both}	$ C $	IEC_{σ}	IEC_{max}	IEC_{rate}	IEC_{both}
500	10	5012	50.4	375	0.0748	4637	4994	50.2	379	0.0759	4615
	20	5012	50.4	375	0.0748	4637					
	50	5012	50.4	375	0.0748	4637					
	100	5012	50.4	375	0.0748	4637					
1000	10	50792	374.0	1661	0.0327	49131	50779	377.7	1616	0.0318	49163
	20	50792	374.0	1661	0.0327	49131					
	50	50792	374.0	1660	0.0327	49132					
	100	50792	374.0	1660	0.0327	49132					
2000	10	97321	376.7	1399	0.0144	95922	97320	377.9	1397	0.0144	95923
	20	97321	376.7	1399	0.0144	95922					
	50	97321	376.7	1399	0.0144	95922					
	100	97321	376.7	1399	0.0144	95922					
978(actual)	10	54851	195.8	1732	0.0316	53119	54800	204.6	1683	0.0307	53117
	20	54851	195.8	1732	0.0316	53119					
	50	54851	195.8	1732	0.0316	53119					
	100	54851	195.8	1732	0.0316	53119					

成されたテストの内、露出数の総和が小さいテストを多く等質テストに含めることができていることから確認できる。

5.4 $OC = 5$ での等質テストの列挙数 U_v の最適化

$OC=5$ での、等質テストの生成における列挙数 U_v を 5,10,20,100 としたときに選択された等質テストを以下の表 10 に示す。テストの構成には植野らの手法 [17] を用いた。最大クリーク探索の時間は 2 時間で行った。

この結果より、アイテムバンクサイズが 1000 以外の場合で等質テストの構成数と最大露出数の差は提案手法の方が大きくなり、等質テスト構成数と最大露出数の両方を考慮した時に改善できた。また、探索時間が等しい場合であっても隣接したテストからなる部分グラフに対して最大クリーク探索を行うことで構成数が大きい等質テストを構成することができることがわかった。

6 むすび

本論文では、植野らの手法でテスト構成を行い、露出数の総和順に総和の小さいテストを必ず含むように部分グラフの抽出を行い、部分グラフに対して最大クリーク探索を行うことで複数の等質テストを構成し、それらの等質テストを比較する手法を提案した。また、列挙された等質テストの中から最大露出数とテスト構成数の両方を考慮した等質テストを選択した。本手法では $OC = 0$ のような疎なグラフに対

しては部分グラフを抽出することでクリーク探索の時間を削減できることがわかった。しかし、 $OC = 5$ のような最大クリークが多数得られるようなグラフに対しては抽出の際にテストを総和の順に選択し、そのテストに隣接するような部分グラフを抽出しているので、並べ替えの方法によっては効果的な部分グラフを得られるまでに何度も繰り返す必要があり、時間がかかるという問題がある。そこで、最大露出数が小さくなるような等質テストに含まれるテストの傾向を調べることで、テストの並べ替え方を変更し、この問題を解決できると考える。また、整数計画法を用いた手法への適用を行っていく。

参考文献

- [1] 植野真臣 and 永岡慶三,e テスティング, 培風館,2009.
- [2] Maomi Ueno, Kazuma Fuchimoto, and Emiko Tsutsumi:E-testing from artificial intelligence approach. Behaviormetrika, Vol. 48, No. 2, pp. 409–424, 2021.
- [3] Wim J van der Linden. Liner Models for Optimal Test Design. Springer, 2005.
- [4] Koun Tem Sun, Yu Jen Chen, Shu Yen Tsai, and Chien Fen Cheng. Creating irt-based parallel test forms using the genetic algorithm method. Applied measurement in education, 21(2):141–161, 2008.
- [5] Pokpong Songmuang and Maomi Ueno. Bees algorithm for construction of multiple test forms in e-testing. IEEE Transactions on Learning Technologies, 4(3):209–221, 2010.
- [6] Dmitry I Belov and Ronald D Armstrong. A constraint programming approach to extract the maximum number of non-overlapping test forms. Computational Optimization and Applications, 33(2-3):319–332, 2006.
- [7] 石井隆稔, 植野真臣, et al. 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. 電子情報通信学会論文誌 D, 97(2):270–280, 2014.
- [8] Takatoshi Ishii, Pokpong Songmuang, and Maomi Ueno. Maximum clique algorithm and its approximation for uniform test form assembly. IEEE Transactions on Learning Technologies, 7(1):83–95, 2014.
- [9] Etsuji Tomita, Sora Matsuzaki, Atsuki Nagao, Hiro Ito, and Mitsuo Wakatsuki. A much faster algorithm for finding a maximum clique with computational experiments. Journal of Information Processing, 25:667–677, 2017.
- [10] Chu Min Li, Hua Jiang, and Felip Many ‘ a. On minimization of the number of branches in branch-and-bound algorithms for the maximum clique problem. Computers & Operations Research, 84:1–15, 2017.

- [11] 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, pages 102–112. Springer, 2017.
- [12] 石井隆稔, 赤倉貴子, and 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. 電子情報通信学会論文誌 D, 100(1):47–59, 2017.
- [13] 渕本壱真, 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, 電子情報通信学会論文誌 D, Vol.J103-D, No.12, pp.881-893 (2020)
- [14] Howard Wainer. Cats: Whither and whence. *Psicologica*, 21(1):121–133, 2000.
- [15] Takatoshi Ishii and Maomi Ueno. Clique algorithm to minimize item exposure for uniform test forms assembly. In International Conference on Artificial Intelligence in Education, pages 638–641. Springer, 2015.
- [16] T. Ishii and M. Ueno, “Clique algorithm to minimize item exposure for uniform test forms assembly,” International Conference on Artificial Intelligence in Education Springer, pp.638–641 2015.
- [17] 植野晶, 渕本壱真, 植野真臣. 項目露出を考慮した整数計画法による等質テスト構成. 2021 年度人工知能学会全国大会
- [18] Ronald D Armstrong, Douglas H Jones, and Zhaobo Wang. Automated parallel test construction using classical test theory. *Journal of Educational Statistics*, 19(1):73–90, 1994.
- [19] Wim J van der Linden and Jos J Adema. Simultaneous assembly of multiple test forms. *Journal of educational measurement*, 35(3):185–198, 1998.
- [20] Frederic M Lord and Melvin R Novick. Statistical theories of mental test scores. IAP, 2008.
- [21] Frank B Baker and Seock Ho Kim. Item response theory: Parameter estimation techniques. CRC Press, 2004.
- [22] Ronald D Armstrong, Douglas H Jones, and Charles S Kunc. Irt test assembly using network-flow programming. *Applied Psychological Measurement*, 22(3):237–247, 1998.
- [23] Ellen Boekkooi-Timminga. The construction of parallel tests from irt-based item banks. *Journal of Educational Statistics*, 15(2):129–145, 1990.
- [24] Panos M. Pardalos and Jue Xue. The Maximum Clique Problem. *Journal of Global Optimization* volume 4, pages 301–328 ,1994

- [25] Maomi Ueno, Kazuma Fuchimoto, and Emiko Tsutsumi: E-testing from artificial intelligence approach. *Behaviormetrika*, Vol. 48, No. 2, pp. 409–424, 2021.