eテスティングの発展と課題

植野真臣
電気通信大学大学院
情報理工学研究科

本研究は　基盤研究(S) 19H05663 「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」（研究代表者: 植野真臣）の助成を受けている.

本講演の内容

eテスティングの発展と課題について紹介する.

1. eテスティングにおける等質テスト生成の最先端技術：45万の等質テスト生成に成功した！！
2. 問題項目の露出率の偏りを軽減するための等質適応型テストの最先端手法
3. 露出率の偏りを軽減する等質テスト生成最先端アルゴリズム

eテスティングの基礎およびこれまでの研究

- 科研費基盤研究（S）シンポジウム 「eテスティング最前線」（科学研究費基盤研究(S) 19H05663（研究代表者：植野真臣））の以下より　資料をダウンロードしてください。

- http://www.ai.lab.uec.ac.jp/kakenhi/

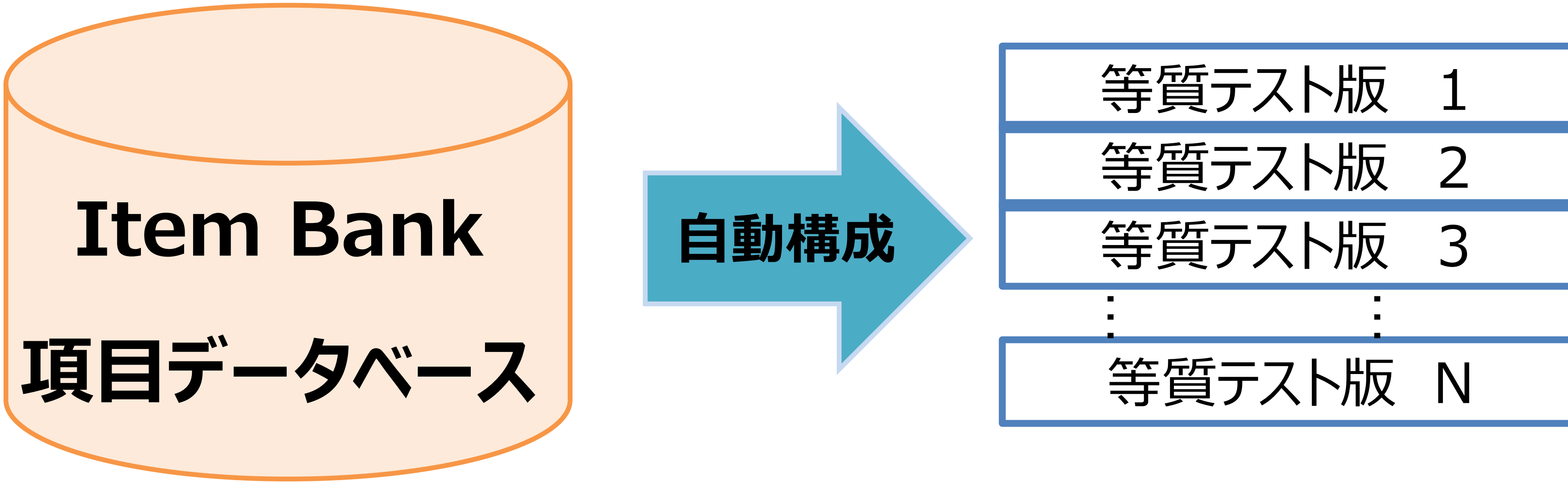
- 初心者の方には
- eテスティングの基礎を　まず読んでいただくことをお勧めします。

eテスティング

- 同一能力の受検者が異なるテストを何度受検しても同一の得点スコアを返すCBT
- 信頼性の高いテストを実現する技術
- 心理学×統計×コンピュータサイエンスの複合分野

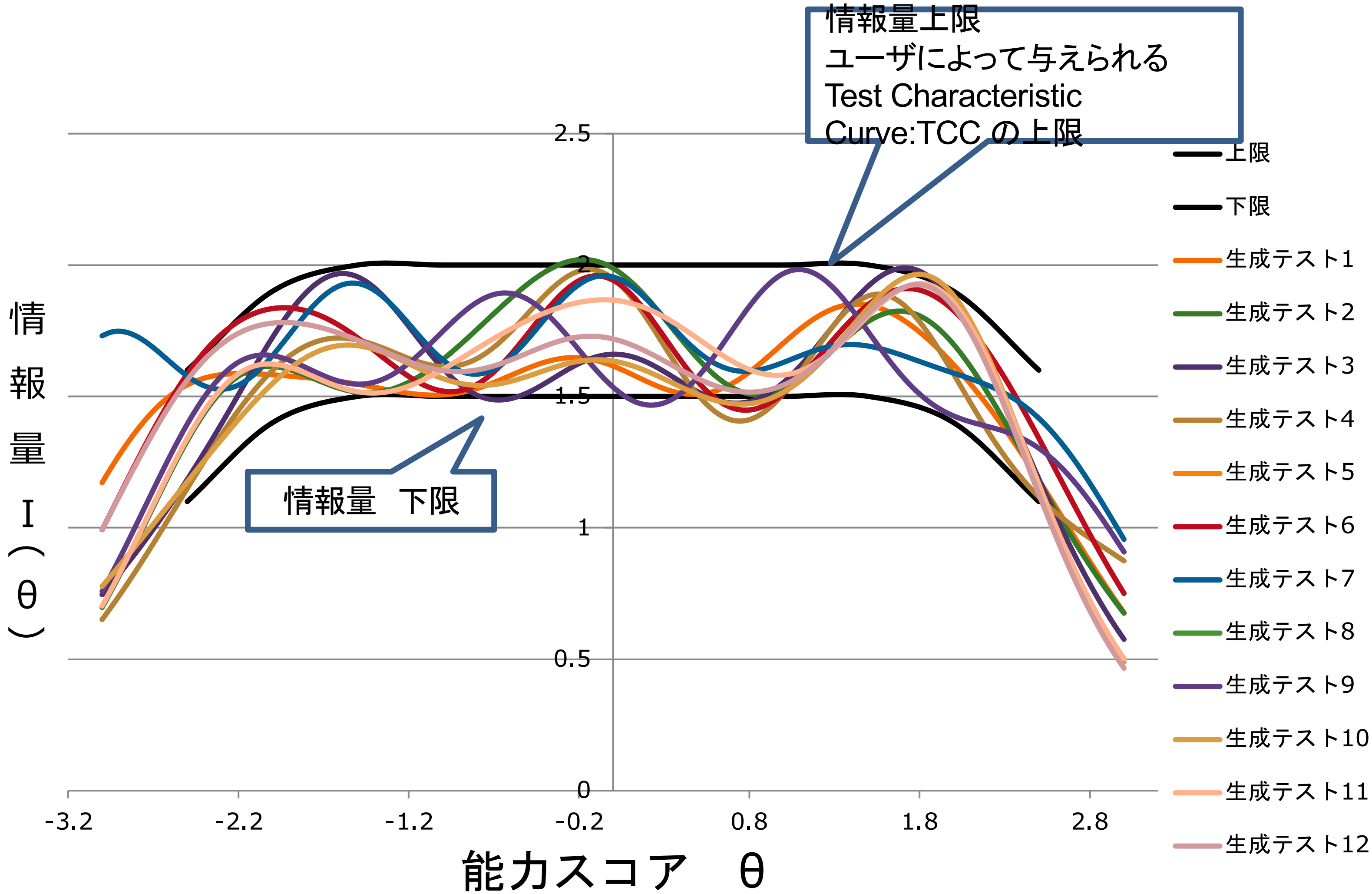
eテスティング：等質テスト版生成

できるだけ多くの等質テスト版を生成

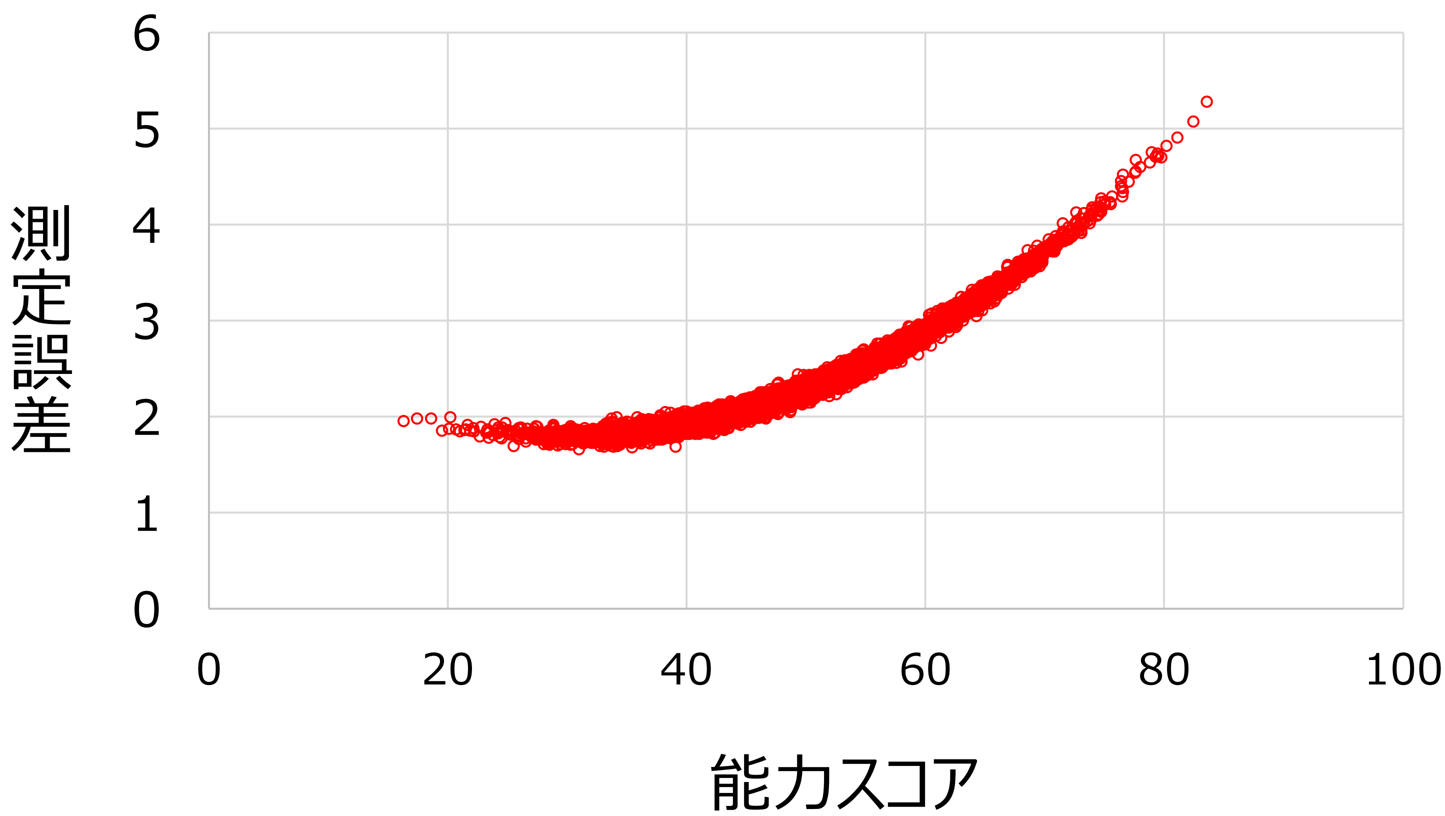


情報処理技術者試験など我が国のeテスティングでは
２００００問程度の問題項目がアイテムバンクに構築
されていることが多い．　毎年項目は追加されるが，
作問した問題項目がすべて登録されるわけではない．

テスト版の情報量を等質に自動生成する



ある公的試験の実際の等質テストのスコアの測定誤差



どのように等質テストの項目組み合わせを求めるのか？

①Big Shadow Test

W. J. van der Linden, : Liner Models for Optimal Test Design, Springer (2005)

線形計画法を用いた計算量の小さい実用的手法。欠点として生成されるテストの誤差が徐々に大きくなる。

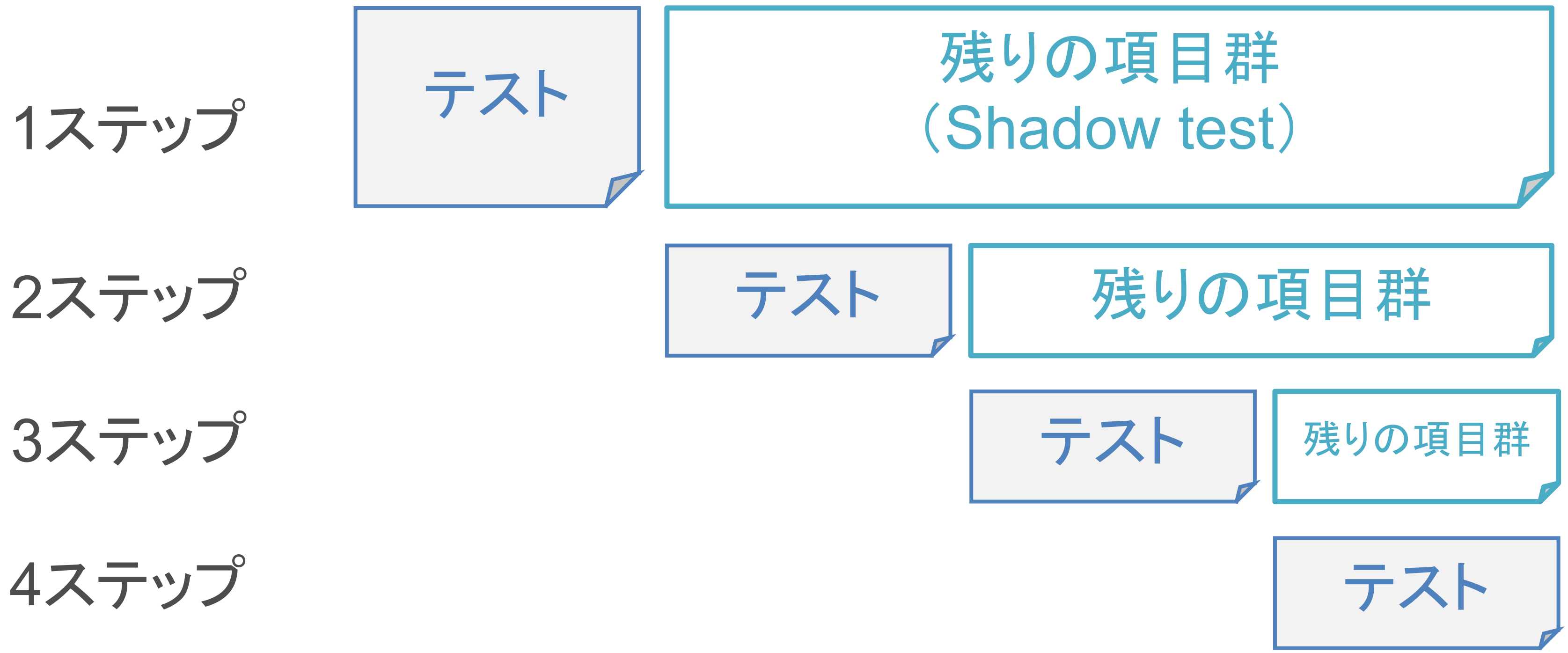
②最大クリーク手法

Takatoshi Ishii, Pokpong Songmuang, Maomi Ueno, "Maximum Clique Algorithm and its approximation for Uniform Test Form Assembly", IEEE Transactions on Learning Technologies, Vol.7(1), pp.83-95, 2014.

最大クリーク問題として解く。計算量は大きいが、厳密解が求められる。

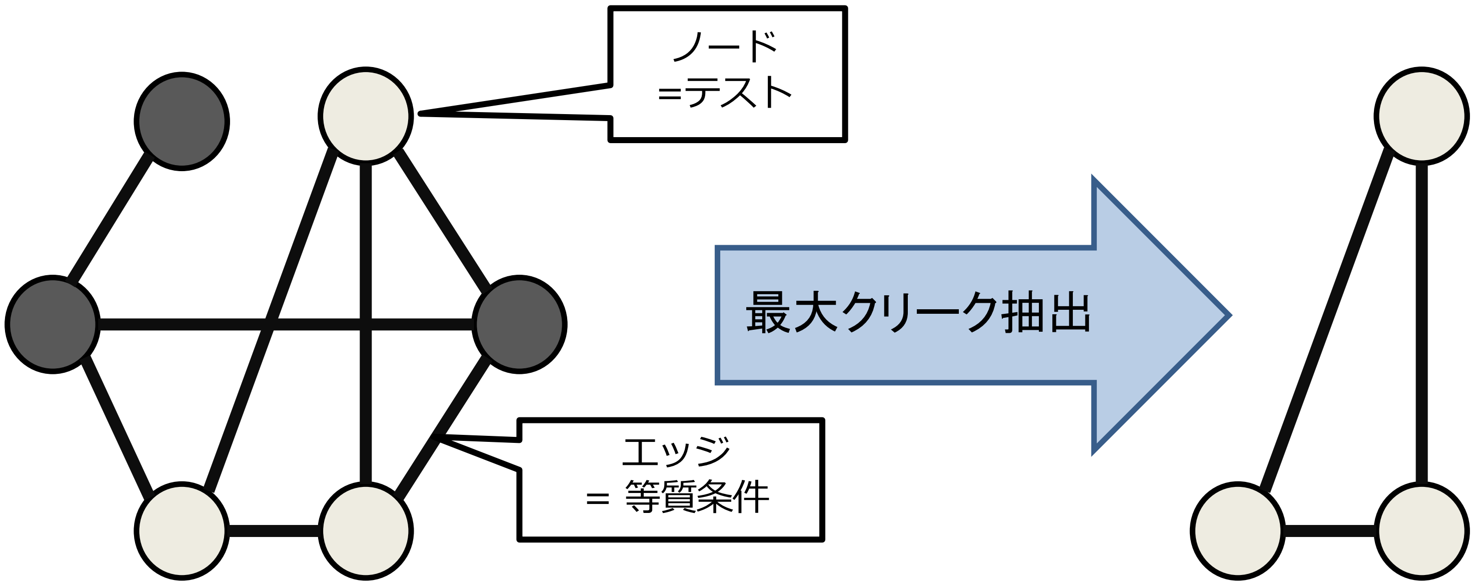
①Big shadow test(W. J. van der Linden2005)

1. アイテムバンクから逐次的にテストを構成
2. 「構成中のテスト」と「アイテムバンクの残りの項目群の平均」のテスト情報量の差異を線形計画法により最小化



②最大クリーク法（現在 最も正確で最大のテスト構成数を可能にする方法)

2018年度電子情報通信学会論文賞受賞



1. Takatoshi Ishii, Maomi Ueno, “Algorithm for Uniform Test Assembly Using a Maximum Clique Problem and Integer Programming”, International Conference on Artificial Intelligence in Education (AIED), LNAI 10331, pp. 102–112. 2017
2. Takatoshi Ishii, Pokpong Songmuang, Maomi Ueno, “Maximum Clique Algorithm and its approximation for Uniform Test Form Assembly”, IEEE Transactions on Learning Technologies, Vol.7(1), pp.83-95, 2014.

アイテムバンク500項目でのテスト構成数

重複項目数	Big Shadow法	最大クリーク法
0	15	10
1	19	22
2	19	65
3	19	223
4	19	936
5	19	4324
6	19	19817
7	19	61740
8	19	93678
9	19	99469
10	19	99979

アイテムバンク1000項目でのテスト構成数

重複項目数	Big Shadow法	最大クリーク法
0	25	17
1	39	61
2	39	282
3	39	1585
4	39	9793
5	39	46162
6	39	90127
7	39	99396
8	39	99979
9	39	99998
10	39	100000

アイテムバンク2000項目でのテスト構成数

重複項目数	Big Shadow法	最大クリーク法
0	61	32
1	79	186
2	79	1436
3	79	12456
4	79	62424
5	79	96859
6	79	99891
7	79	99991
8	79	100000
9	79	100000
10	79	100000

等質テストで何ができるか？

- 毎回の異なるテストが等質であることが保障され、いつでも何度でもテストを受けることができる。
- テスト実施後にその精度が評価でき、テストの評価をその都度行なえる。
- １００点満点の素点を受検者に返せ、それをもとに合否やランキングを決めることができる。

最大クリーク法の問題

- 空間計算量によるテスト構成数の上限
  - 最先端の最大クリークアルゴリズムは MCT [Tomita et al. 2017], MoMC [CM Li, Jiang, and Many' a. 2017]など
  - 時間計算量$O(|2|^{0.19171|V|})$ で探索可能.
  - しかし，空間計算量が$O(|V|^2)$と大きく，最大で10万程度のテスト構成が限界

1. eテスティングにおける等質テスト生成の最先端技術

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

植野研究室　渕本 壱真 (M2)

提案手法：

二段階探索アルゴリズム

[第一段階]

時間計算量$O(|2|^{0.19171|V|})$が小さく，空間計算量$O(|V|^2)$が大きい最大クリーク法で　メモリが許す限りテスト生成

[第二段階]

時間計算量$O(|V|・2^n)$が大きく，空間計算量$O(|V|)$,が小さい整数計画法を用いた手法に切り替えてテスト生成、　並列探索アルゴリズムに改善

淵本壱真，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)

実験

Item bank size	OC	Hybrid RBP	Proposal
500	0	17	17
	5	9,774	14,331
	10	126,987	127,149
1000	0	33	34
	5	62,965	97,492
	10	127,432	131,300
2000	0	69	70
	5	103,164	129,257
	10	126,532	140,700
978	0	35	35
(actual)	5	61,859	73,693
	10	122,681	124,200

→テスト構成数が大きくなると，差異が大きい

実験

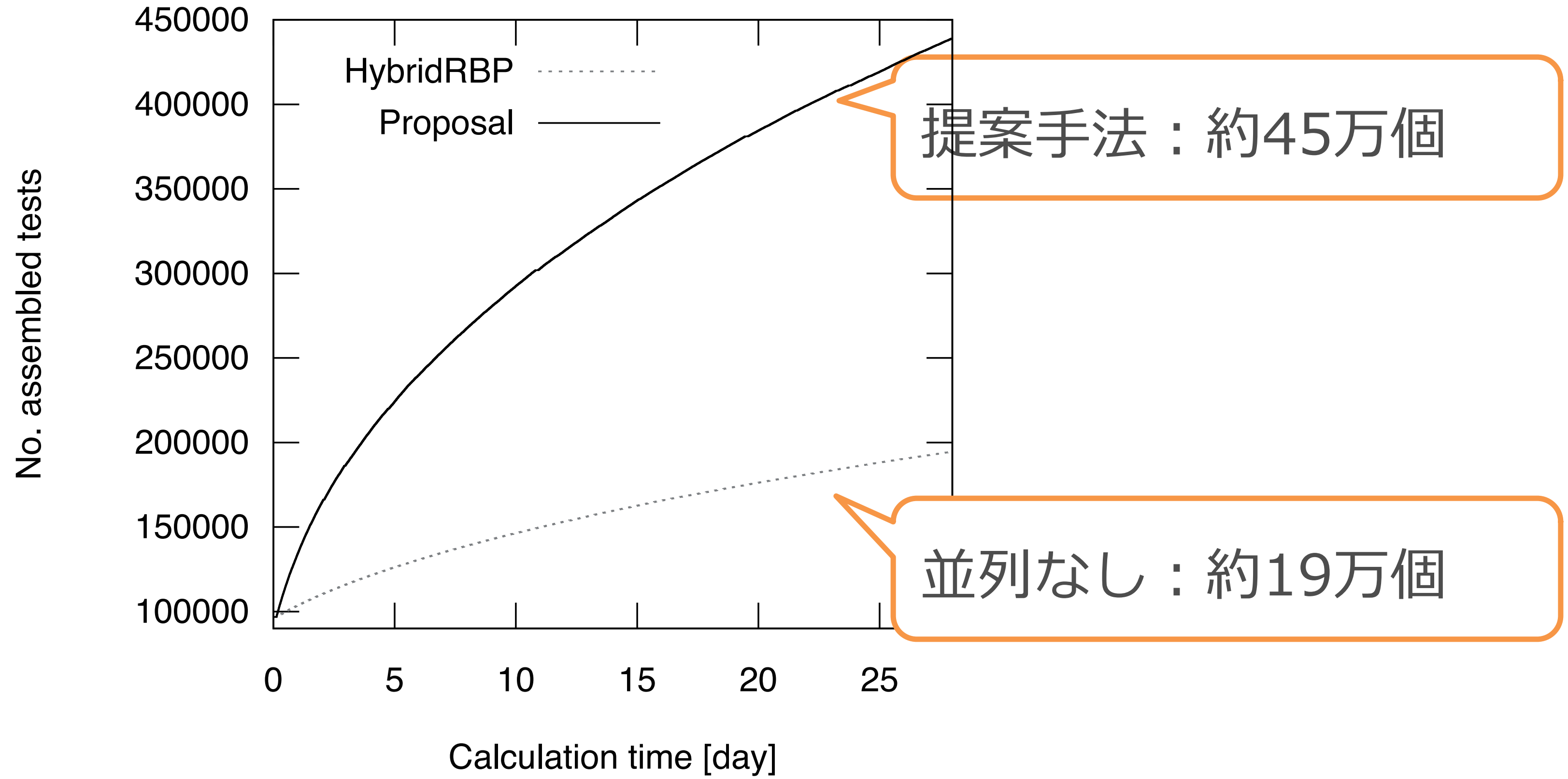
Item bank size	OC	Hybrid RBP	Proposal
500	0	17	17
	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]
1000	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	97,492
	[illegible]	127,432	131,300
2000	[illegible]	69	70
	5	103,164	129,257
	10	126,532	140,700
978	0	35	35
(actual)	5	61,859	73,693
	10	122,681	124,200

大規模アイテムバンクを用いて，

計算時間を28日間まで延長し，追加実験

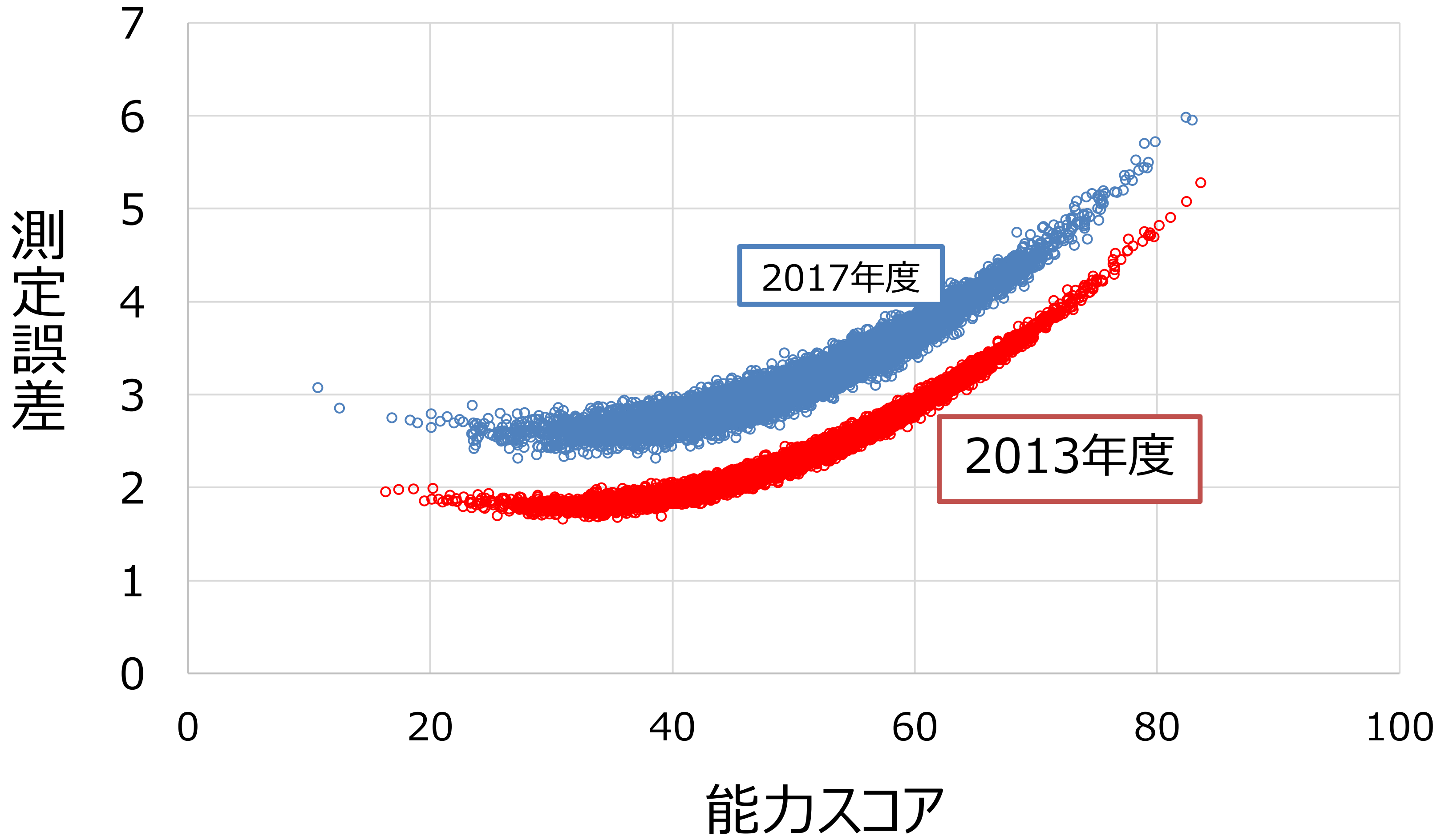
→テスト構成数が大きくなると，差異が大きい

実験



- 約45万の等質テストを生成できた.

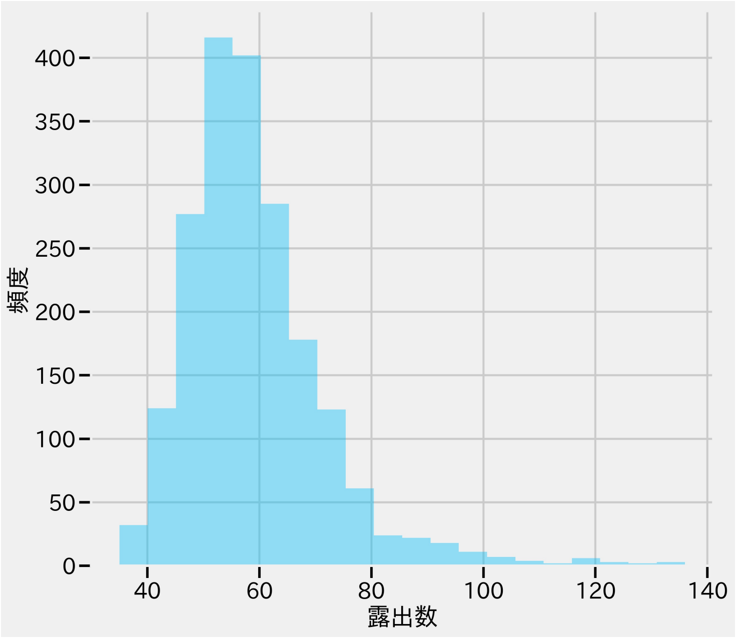
あるハイステークス試験における2013年度と2017年度の実際の等質テストのスコアの測定誤差



項目露出数（出題数）のヒストグラム

- 特定の良問を集中的に選択してしまうことで良問の露出数（出題数）が増加
- 項目の特性が良問から大きく変化

測定誤差の劣化が急速に進行



新しい問題

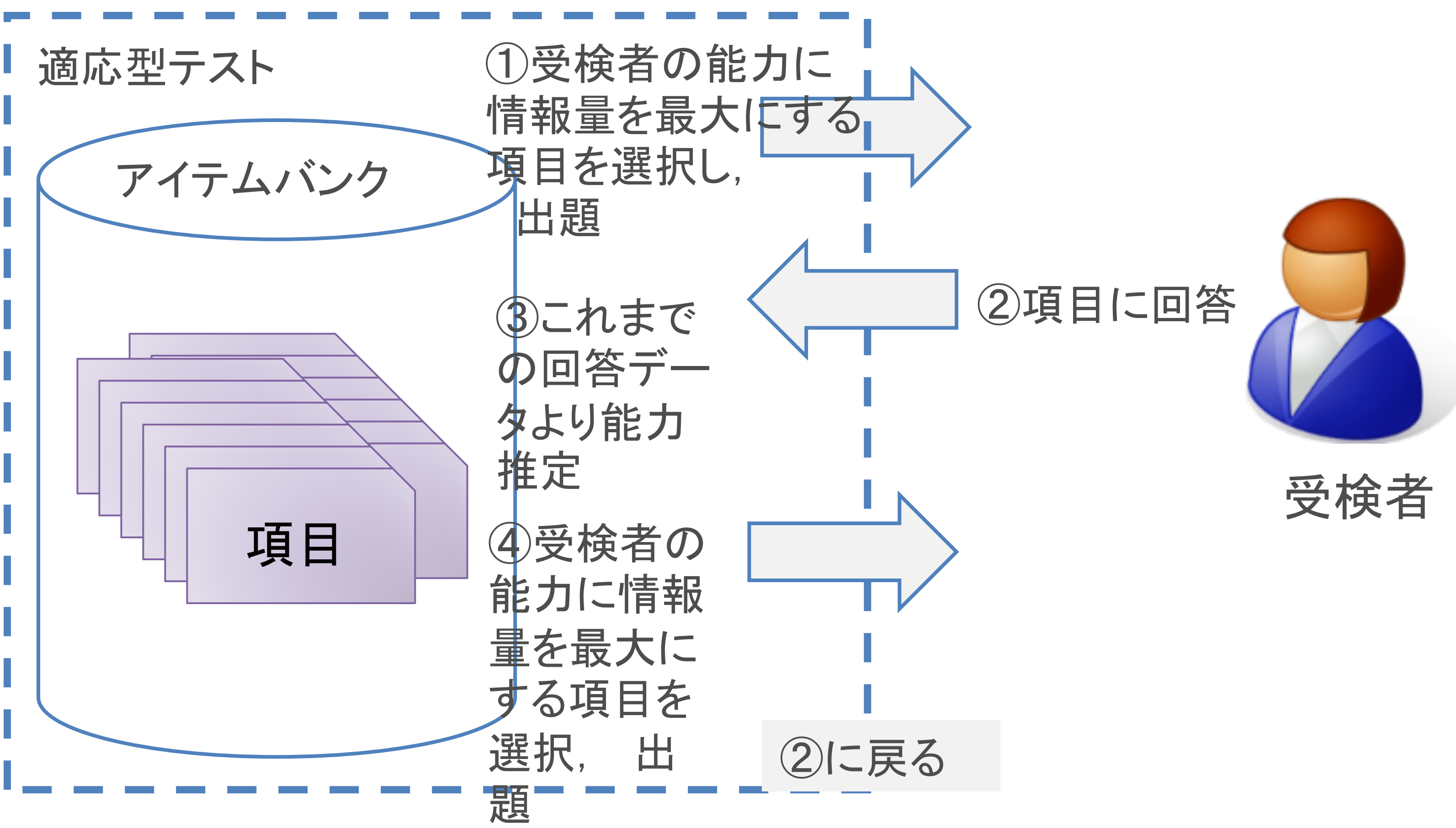
問題項目の露出数の偏りを減少させなければならない！！

問題項目の露出率の偏りを軽減するための等質適応型テストの最先端手法

高精度能力推定を保証する二段階等質適応型テスト

宮澤　芳光（大学入試センター),植野真臣

適応型テストのアルゴリズム



利点

測定精度を下げずに　受検者に出題する問題項目数を減少させることができる

→

問題項目露出数を減少させることができるかもしれない

問題

- 適応型テストは、同一の能力を持つ受検者に同一の問題項目を出題してしまう．　項目露出数の偏りはより増大してしまう．

先行研究

シャドーテストを用いた制約付き適応型テスト（van der Linden, 1998 ）

アイテムバンク内の項目の露出数を制限し,

情報量が最大の項目群（シャドーテスト）を構成

構成された項目群（シャドーテスト）から情報量が最大の項目を選択

アイテムバンク分割法（Kingsbury1989,Hetter1997）

アイテムバンクを複数のグループに分割し,

最も項目の露出数が少ないグループから情報量が高い項目を選択

適格確率を用いた適応型テスト（van der Linden and Veldkam, 2004, 2019）

露出率に基づく適格確率に従って項目集合を構成し,

その項目集合から最も高い情報量をもつ項目を選択

露出数を最小化する適応型テスト（Miyazawa and Ueno, 2019）

情報量を制約条件として露出数が最小の項目集合を逐次的に構成し,

その項目集合から情報量が最大の項目を選択

先行研究

未出題の項目が多いときは，

情報量が高い項目が抽出され，少ない項目数で能力を推定

情報量が高い項目が既に最大露出数に達したときは，

少ない項目数で能力を推定が難しい

→受検者間でテストの長さや測定精度に偏りが生じており，
露出数の減少と測定精度の向上にはトレードオフの関係がある

上記の問題を解決する手法

露出数と測定精度のバランスを制御する等質適応型テスト

（宮澤&植野2018，Ueno and Miyazawa, 2019）

等質な測定精度を持つ等質項目集合を受検者に割り当て，

その等質項目集合から情報量が高い項目を選択する

→しかし，漸近誤差は小さくなるが能力推定値の真値との誤差が大きいことがわかった．

本研究の目的

二段階等質適応型テストを提案

- テストの前半に
  - 等質な部分項目集合から情報量が高い項目を選択し，
- 受検者の能力推定値が収束してきたテストの後半に
  - アイテムバンクからその能力推定値付近で
  - フィッシャー情報量が高い項目を選択

提案手法の特徴

第一段階では,

等質な項目集合の項目難易度分布が能力分布全般に

一様に分布しており，過学習を避け,

高速に推定値が真の能力値近傍まで収束

第二段階では,

アイテムバンク全体からの項目選択で高精度な推定

真の能力値近傍からの項目選択で露出分布の偏りが減少

評価実験

提案手法の有効性を検証するために

実データのアイテムバンク((株)リクルートで開発)を用いて以下の統計量を比較

- アイテムバンク内の項目の露出数
- 受検者に出題されたテストの長さ
- 能力推定値のRMSE

比較対象の手法

- 従来の適応型テスト（CAT）
- 制約付き適応型テスト（van der Linden, 1998; LPと呼ぶ）
- 適格確率を用いた適応型テスト
  （van der Linden and Veldkam, 2004, 2019; LVと呼ぶ）
- アイテムバンク分割法（Kingsbury and Zara, 1989; KZと呼ぶ）
- 露出数を最小化する適応型テスト（Miyazawa and Ueno, 2019; MUと呼ぶ）
- 提案手法（Proposal）

シミュレーションの手続き

1. 受検者の真の能力値を$\theta \sim N(0,1)$ からサンプリング
2. 各手法に応じてアイテムバンクから項目を出題

   （ただし，最大露出数を揃えるため露出数が100以下の項目から選択）
3. 反応データを能力真値と項目パラメータを所与として発生
   - 乱数$U(0,1)$を生成し，乱数の値が選択された項目の正答確率より低ければ正答，それ以外を誤答
4. 解答履歴データから能力推定値$\hat{\theta}$をEAP法により推定
5. テスト終了条件を満たせば終了，そうでなければ2へ

＊ 各手法ごとに1000回反復

実験結果

		テストの長さ		RMSE	露出数	
テスト終了条件	手法	平均	標準偏差		平均	標準偏差
出題項目数が 30 項目	CAT	30	0	0.34	78.95	35.39
	IP	30	0	0.34	79.79	34.76
	LV	30	0	0.33	80.65	33.71
	MU	30	0	0.43	30.67	1.44
	KZ	30	0	0.35	76.53	35.35
	Proposal	30	0	0.32	46.01	43.95

新たな課題

等質適応型テストでは、最大クリーク法(Ishii 他　2014)を用いて　等質グループを生成している。

しかし、そもそも最大クリークが露出分布の偏りを持つ。

等質テスト生成法そのもので　露出分布の偏りがないようにする手法を開発しなければならない.

3. 露出率の偏りを軽減する等質テスト生成最先端アルゴリズム

項目露出を考慮した数計画法による等質テスト構成

電気通信大学大学院情報理工学研究科
植野研究室 植野晶(M1)

研究目的

<u>項目露出数の偏りを減少させながらテスト構成数の最大化</u>

- 項目(問題)の露出数とは
  - 同一項目が等質テスト全体で出題される回数のこと
- 露出数が多いと受験者間で項目が共有される
  - テストの信頼性低下につながる[Wainer 2000]
- テスト構成を保存、露出数の偏りが最も小さい構成を出力[Ishii 2015]

Howard Wainer. Cats: Whither and whence. Psicologica, 21(1):121–133, 2000.
Takatoshi Ishii and Maomi Ueno: Clique Algorithm to Minimize Item Exposure for Uniform Test Forms Assembly, Artificial Intelligence in Education – 17th International Conference, AIED 2015, 638-641

項目露出数を軽減する手法 [Ishii 2015]

先行研究 [Ishii2015]

- 項目iの露出数

$$\sum_{r=1}^{|C|} X_{i,r}$$

C: 構成した等質テスト(クリーク)

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

- 露出数とは同一項目がテスト全体で何回出題されるか

Takatoshi Ishii and Maomi Ueno: Clique Algorithm to Minimize Item Exposure for Uniform Test Forms Assembly, Artificial Intelligence in Education – 17th International Conference, AIED 2015, 638-641

項目露出を軽減する手法 [Ishii 2015]

- テスト構成Cにおける最大露出数

$$E_C = max_{i=1}^{n}(\sum_{r=1}^{|C|} X_{i,r})$$

- 露出率

$$\frac{E_C}{|C|}$$

- [Ishii 2015]では逐次的に等質テストを構成し今までに構成した等質テストを全て保存
  - 最後に露出率が最も低いテスト構成を出力
    - 従来手法(RndMCP等)よりも露出率が低くなった

- MCALIE法と呼ぶ

Takatoshi Ishii and Maomi Ueno: Clique Algorithm to Minimize Item Exposure for Uniform Test Forms Assembly, Artificial Intelligence in Education – 17th International Conference, AIED 2015, 638-641

研究目的

<u>項目露出数の偏りを減少させながらテスト構成数の最大化</u>

- [Ishii 2015]で構成したテストでは露出数の偏りが生じてしまう

- 本研究の目的
露出数の偏りを軽減しながら<u>テスト構成数の最大化</u>

Takatoshi Ishii and Maomi Ueno: Clique Algorithm to Minimize Item Exposure for Uniform Test Forms Assembly, Artificial Intelligence in Education – 17th International Conference, AIED 2015, 638-641

[提案手法]整数計画法を用いた露出数の偏りを軽減する等質テスト構成

<u>提案1</u> 整数計画法を用いた頂点生成

- 制約を満たす頂点を一様に選択
  - 露出数の偏りを改善

ただし整数計画法を用いても、項目露出数の一様性は考慮していないので、依然露出数が大きい項目が存在

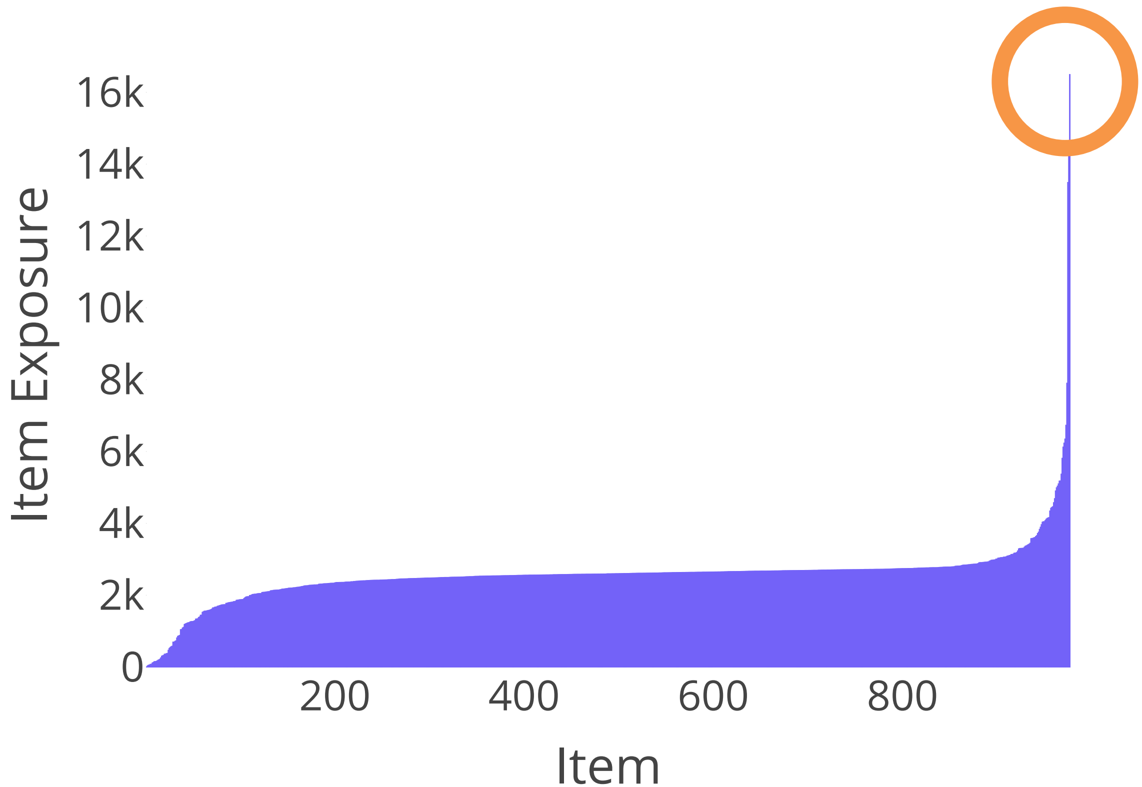
<u>提案2</u> 露出数上位s位までの項目を除外したアイテムバンクから頂点生成

- 露出数上位以外の項目を組み合わせて頂点生成
  - 最大露出数を抑制
  - 特定の項目が過度に出題されない

シミュレーション・実データより，露出数の偏りを軽減し，さらにテスト構成数も増加できたことが示された.

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，電子情報通信学会論文誌 D (2021)

MCALIE法における特定項目の過度な出題



$n = 978$(実データ)
$OC = 10$

- ➢ MCALIE法で構成したテストの項目を露出数の昇順にソート

テストの構成数: 102,862
露出数の最大値: 16,698

- ➢ 16%以上のテストに特定の項目が出題されてしまう

整数計画法による頂点生成[提案手法]

提案2 露出数上位s位までの項目を除外した
アイテムバンクから頂点生成

- 露出数上位以外の項目を組み合わせて頂点生成
  - 最大露出数を抑制
  - 特定の項目が過度に出題されない

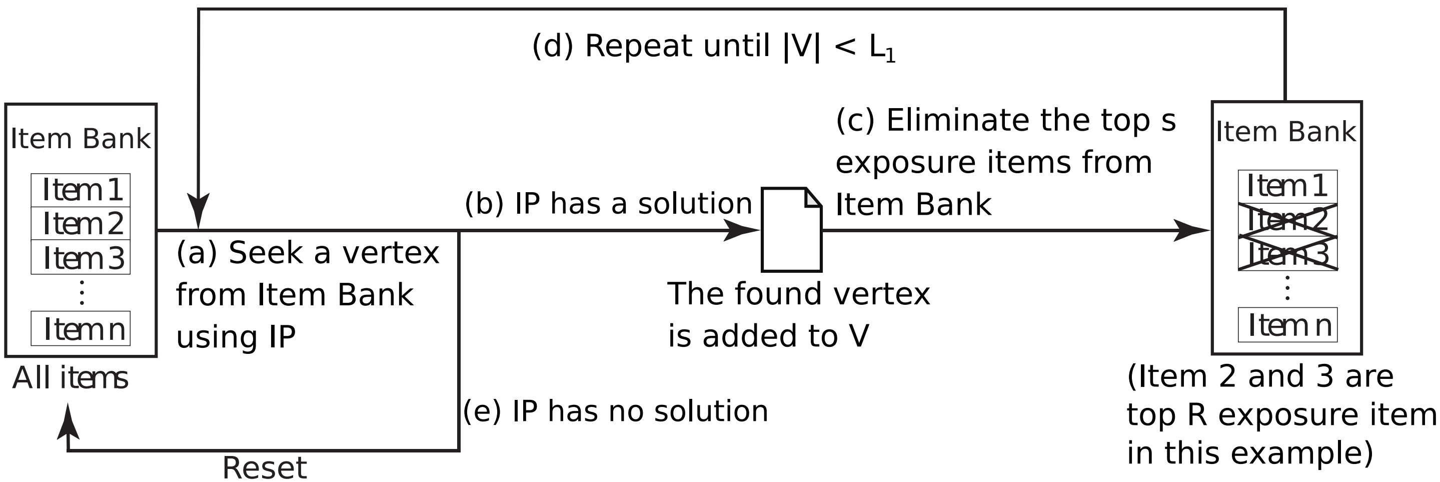
植野晶, 渕本壱真, 植野 真臣：項目露出を考慮した整数計画法による等質テスト構成,
電子情報通信学会論文誌 D (2021)

整数計画法による頂点生成[提案手法]

頂点生成時に項目露出の偏りを小さくする理由

- 頂点生成後はグラフを構築し、最大クリーク探索
  - 項目露出の偏りが小さいグラフ(頂点集合)において最大クリーク探索
  - グラフ探索後のクリーク(等質テスト)の項目露出の偏りも小さくなる

頂点生成アルゴリズム[提案手法]



a. 整数計画法でアイテムバンクから頂点を生成
b. 解(頂点)が存在した場合は頂点集合Vに追加
c. 頂点集合V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
d. 頂点集合Vの大きさがL_1になるまでa~cを繰り返す
e. 整数計画問題の解が存在しない場合は除外した項目をアイテムバンクに戻す

整数計画問題の定式化[提案手法]

条件を満たすテスト集合の中から乱数で等質テストを生成し、露出数の偏りを軽減

$\boldsymbol{variable}$

$$x_i = \begin{cases} 1 & (\text{項目}i\text{を頂点に選ぶ}) \\ 0 & (\text{項目}i\text{を頂点に選ばない}) \end{cases}$$

$\boldsymbol{maxmize}$

$$\sum_{i=1}^{n} \lambda_i x_i$$

λ: 0～1の一様乱数

$\boldsymbol{subject\ to}$

$$\sum_{i=1}^{n} x_i = M$$

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

$$(\mathrm{k} = 1, \ldots, \mathrm{K})$$

整数計画法による項点生成の実験[提案1]

- 整数計画法を用い項点生成すると露出数の偏りが抑制されることをRndMCP法との比較実験で確認

- 以下の条件で実験を行なった

アイテムバンク:

シミュレーション($n = 500, 1000, 2000$)

実データ($n = 978$)

テスト項目数: $M = 25$

テスト情報量の下限/上限:

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

重複項目数: $OC = \{0, 5, 10\}$

整数計画法による項点生成の実験結果[提案1]

Pool Size (n)	OC	RndMCP テスト構成数	RndMCP 露出数の標準偏差	RndMCP 最大露出数	RndMCP 露出率	整数計画法 テスト構成数	整数計画法 露出数の標準偏差	整数計画法 最大露出数	整数計画法 露出率
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 (実データ)	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%**

- 整数計画法を用いることで露出数の偏りが抑えられ露出率が改善
- 整数計画法を用いるだけでは,
  $n = 1000, 978$のアイテムバンクで依然として
  特定の項目が4%以上のテストに出題されている

統合比較実験の結果

Pool Size (n)	OC	RndMCP/MCALIE			RndMCP/HybridRBP			提案手法/MCALIE			提案手法/HybridRBP		
		テスト構成数	最大露出数	露出率	テスト構成数	最大露出数	露出率	テスト構成数	最大露出数	露出率	テスト構成数	最大露出数	露出率
500	0	**18**	1	**5.6%**	**18**	1	**5.6%**	**18**	1	**5.6%**	17	1	5.9%
	5	11475	858	7.5%	15795	1216	7.7%	13430	998	**7.4%**	**16016**	1192	7.4%
	10	**104914**	13267	12.6%	104817	13332	12.7%	99970	5370	**5.4%**	104684	5720	5.5%
1000	0	**36**	1	**2.8%**	35	1	2.9%	35	1	2.9%	34	1	2.9%
	5	55047	3691	6.7%	55087	3712	6.7%	50781	1616	**3.2%**	**59316**	1986	3.3%
	10	**104260**	8762	8.4%	104228	8724	8.4%	99998	2847	**2.8%**	104250	3031	2.9%
2000	0	**70**	1	**1.4%**	69	1	1.4%	**70**	1	**1.4%**	**70**	1	**1.4%**
	5	99620	4028	4.0%	99783	4031	4.0%	97348	1397	**1.4%**	**100196**	1452	1.4%
	10	102775	4105	4.0%	**102849**	4106	4.0%	100000	1418	**1.4%**	102839	1480	1.4%
978 (実データ)	0	**36**	1	**2.8%**	35	1	2.9%	**36**	1	**2.8%**	35	1	2.9%
	5	51644	5342	10.3%	51984	5357	10.3%	54800	1683	**3.1%**	**59668**	1868	3.1%
	10	**102862**	16698	16.2%	102624	16686	16.3%	100000	2754	**2.8%**	102765	2881	2.8%

- 提案手法/MCALIE法の組み合わせ
  - テスト構成数が多くの条件で従来手法よりも減少
- 提案手法/HybridPBR法の組み合わせ
  - テスト構成数: 従来手法とおおよそ同等、$OC = 5$では増加
  - 露出率: 従来手法を大幅に改善

まとめ

- eテスティングの最先端技術を紹介した.

1. eテスティングにおける等質テスト生成の最先端技術：45万の等質テスト生成に成功した！！
2. 問題項目の露出率の偏りを軽減するための等質適応型テストの最先端手法
3. 露出率の偏りを軽減する等質テスト生成最先端アルゴリズム

今後の課題

- 露出率を考慮した等質テスト構成手法を並列化
- 露出率を考慮した等質テスト構成手法を用いた等質適応型テストシステムの開発

おしまい

謝辞
本研究は　基盤研究(S) 19H05663 「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」（研究代表者: 植野真臣）の助成を受けました.