

平成 30 年 6 月 6 日現在

機関番号：32660

研究種目：若手研究(B)

研究期間：2015～2017

課題番号：15K16260

研究課題名(和文) e テスティングの信頼性向上の為に項目曝露を考慮した複数等質テスト構成手法の開発

研究課題名(英文) the developing of uniform test assembly for reducing item exposure and increasing the number of tests

研究代表者

石井 隆稔 (Ishii, Takatoshi)

東京理科大学・工学部情報工学科・助教

研究者番号：40741591

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：本研究は普及が見込まれるeテストの信頼性向上のために、等質な複数のテストをより多く、計算機により構成可能とするための研究である。特にこの課題では、出題される各々の問題(項目)の曝露を低減し、問題流出などのリスクを低減するテスト構成手法の開発を目指した。初年度は、項目曝露を削減するための定式化と単純なアルゴリズムの開発がおおむね完了したが、2,3年度目は研究代表者である石井が体調を崩したため、このアルゴリズムの検証と論文執筆が滞った。二年目にはこの研究の副産物として、より少ない計算コストでより多くのテストを構成可能なアルゴリズムの開発できた。本論文は電子情報通信学会で論文賞を受賞した。

研究成果の概要(英文)：The uniform tests (it also called parallel tests) are used for increasing the reliability of e-testing. In 2007, ISO/IEC23988:2007 was established, and it commends to use the uniform tests for high-stakes test. With the rising of e-testing, uniform tests assembly method is needed to increase the number of constructing tests, and to reduce item exposure, for the preventing testing risks such as item leaking and examinee cheating. In first year, we established the formulation and a naive algorithm for reducing item exposure. In second and third year, by principal investigator got ill, improving the algorithm and writing paper were suffered. In second year, an algorithm for increasing the number of uniform tests was published. This algorithm has been one of spin-offs of the algorithm for reducing item exposure. The published algorithm reduces the cost of calculation, and accelerates constructing uniform tests. The paper of this algorithm received the paper award in IEICE.

研究分野：テスト工学

キーワード：複数等質テスト アルゴリズム 整数計画問題 eテスト

1. 研究開始当初の背景

近年、情報技術の発達により、様々な試験が、計算機とインターネットを用いて作成・配信・実施・採点される、e テスティングが普及しつつある。例えば、情報処理技術者試験や医学部共用試験などのハイステークス・テスト(high-stakes test)においてもe テスティングによる評価の導入・運用が始まっている。ハイステークス・テストとは試験結果の重要性が高い試験であり、このような試験では試験の結果や運用に高い信頼性が求められる。例えば、毎年行われる資格試験などでは経年でテストの難易度が大きく変わってはならない。また試験問題の流出などのリスクに対しても頑健である必要がある。

既に現在、2007年には情報技術を用いた評価の実施に関する国際標準規格(ISO/IEC 23988:2007 Information technology -- A code of practice for the use of information technology (IT) in the delivery of assessments, 対応国内規格 JIS X 7221:2011 アセスメント提供における情報技術(IT)利用の規範)が策定されている。この標準規格のなかでハイステークス・テストでは、異なる項目により構成されているにもかかわらず各テストが同じ能力を同じ精度で測定可能である複数等質テストを使用することが条件として記載されている。日本国内においてはISO及びJISの認定は未だ無いが、情報処理技術者試験ITパスポートや医療系大学間共用試験はこれらの規格を参考に作成されており、複数等質テストにより運用されている。

このような大規模に行われるハイステークス・テストでは可能な限り多くの複数等質テストを作成することが必要となる。

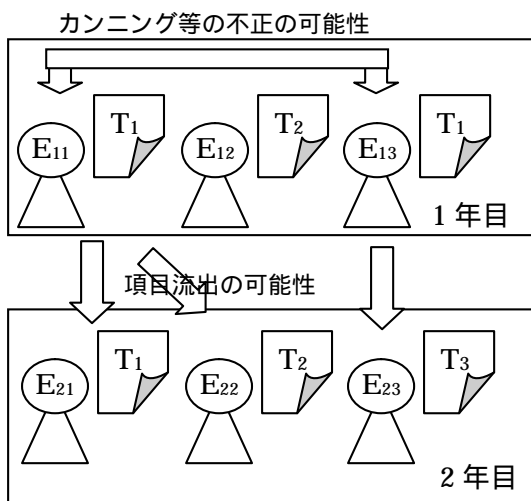


図1 少数の複数等質テストを使用する場合の受験者へのテスト配布の例

例えば図1は経年で行われるハイステークス・テストにおいて少量の複数等質テストを使用する場合に想定されるリスクを示している。1年目においては受験者 E_{11} と E_{13} には同じテストが配布されており、カンニング等

のリスクが生じ得る。また経年を通じて同じテストを使用する場合 (E_{11} と E_{21} など) 過去の出題からテスト問題(以降、項目)が流出してしまう可能性がある。これらの状況は受験者の能力を正しく推定できなくなる為、特にハイステークスなテストでは避けるべき状況となる。そのため、複数等質テストは可能な限り多く作成する必要がある。

近年のe テスティングでは、出題可能な項目のデータベースから出題する項目集合を選び出すことによりテストとする、アイテムバンク方式が採用されることが多い。

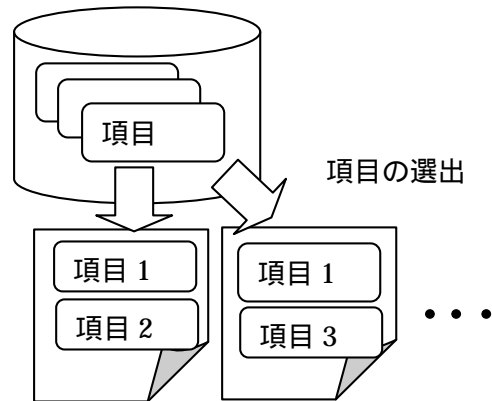


図2 アイテムバンク方式によるテスト構成の模式図

図2はアイテムバンク方式によるテスト構成の模式図である。アイテムバンクには項目そのものだけでなく、例えば正答率や項目反応理論などで導かれるパラメータが保存されている。そのため、それらの項目の組み合わせであるテストの平均点や得点分布などが予測可能となっており、自動テスト構成とは、ユーザによって指定された正答率などのパラメータが予測されたそれに近くなるような項目の組み合わせを探索する最適化問題として定式化される。

このようなアイテムバンク方式での複数等質テストを用いたテストシステムの運用上では、項目露出が問題となる。項目露出とは、それぞれの項目がどれだけ出題されたかであり、図2の例では項目1の露出が項目2、3に比べ多いということになる。通常、複数等質テストは異なる項目を組み合わせ作成することが望ましいが、例えば、どの受験者にも共通で出題すべき項目がある場合や、構成テスト数を多くするために項目のテスト間の項目の重複を許す場合も多い。そのような場合は、(1)テスト間の共有項目数の上限と(2)複数等質テスト内でそれぞれの項目の露出回数管理が必要となる。

2. 研究の目的

研究代表者のこれまでの研究[1-3]では、テスト間の重複項目数については考慮してきたが、構成された複数等質テスト数全体でそれぞれの項目がどれだけ使用されているか、

つまり項目の露出については管理してこなかった。そこで本研究では項目の露出を管理可能な複数等質テスト構成手法の開発を目的とした。

3. 研究の方法

3.1. 項目露出を考慮した複数等質テスト構成の定式化

本研究でもこれまでの研究と同様にテスト構成をクリーク問題として扱った。以下のようなグラフを考えると、複数等質テスト構成は一種のクリーク問題となる。

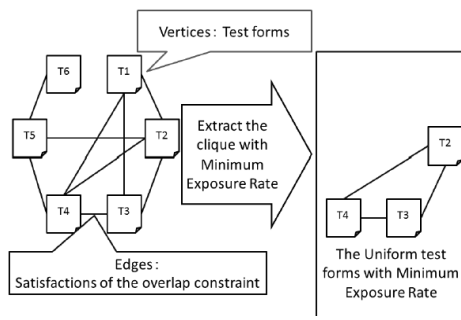


図 3 複数等質テスト構成のためのグラフ構造

グラフの頂点を、重複条件・露出項目条件を除くテスト構成条件を満たすテスト(項目の集合)群(可能テストと呼ぶ)、辺を重複条件の満足としたグラフを考える。

図3は提案手法中で使用されるグラフ構造の例示である。このグラフには6つのテストを表す頂点と重複条件の満足である9つの辺がある。ここで重複条件の満足とは2つのテスト間の共通項目がユーザによって設定される値以下であることを指す。

このようなグラフ中において、クリークと呼ばれる、その頂点集合中のどの2頂点も接続されている頂点集合、はすべてのテストが等質でどの二つのテストも共有項目数が少ないテスト群、つまり複数等質テスト群となる。このようなグラフを以降、複数等質テスト構成のためのグラフ構造と呼ぶ。

項目露出がある値以下の複数等質テストを構成するという事は、このグラフ中なるべくそれぞれの項目が均等に使われているクリークを探索する最適化問題として定式化可能である。

例えば以下のような定式化が可能である。

変数	C, V	(等質テストの集合)
目的関数(最大化)	$ C $	(集合 C の要素集)
制約条件	$u \in C, u$ はテスト構成条件を満たす $u, v \in C,$ $ u, v $ (ユーザが指定する重複項目数) $u \in U_i, i \in u, U_i \in C$ $ U_i $ (ユーザが指定する最大露出回数)	

この定式化では、ユーザによって指定され

た最大露出回数を超えない範囲で、可能な限り多くの複数等質テストを構成する。

また以下のような定式化ではユーザによって指定されたテスト数を構成その中でそれぞれの項目露出回数を可能な限り低減する。

変数	C, V	(等質テストの集合)
目的関数(最小化)	$\text{Max}(U_i)$	(最も使用されている項目を含むテスト数)
制約条件	$u \in C, u$ はテスト構成条件を満たす $u, v \in C,$ $ u, v $ (ユーザが指定する重複項目数) $u \in U_i, i \in u, U_i \in C$ $ C $ (ユーザが指定する必要テスト数)	

これらの定式化をテスト構成によって使い分けることで可能な限り露出数を低減したり、限られた露出数の中で最大限多くのテストを構成できる。

3.2. 整数計画問題を用いたアルゴリズムの効率化

ただし、これらの最適化問題を解くためには、基本的には与えられたグラフ中のクリークを順次列挙しそれぞれのクリークを評価していく必要がある。しかし、複数等質テスト構成の為のグラフは非常に頂点数が多くなるため既存のアルゴリズムでは、効率の良い探索が難しかった。これはこのグラフの頂点が項目の組み合わせで有り、与えられたアイテムバンクサイズに対して組み合わせ爆発的に増えるためである。そのため、研究代表者らの従来研究である露出を考慮せずテスト構成数を最大化する手法[1-3]では、これらの頂点(条件を満たすテスト)をランダムに作成し、グラフを作成しそのグラフ中から最大クリークなどの探索を行っていた。ただし、本手法は計算コスト、とりわけ空間計算量が大きく、十分な主記憶装置を持たない計算環境ではテスト構成できない問題があった。

そこで本研究では新たに効率よくこのようなグラフからの探索を行えるアルゴリズムを提案した。

本アルゴリズムを以下に示す。

1. クリーク $C =$ とする
2. クリークリスト $L =$
3. ランダムにテスト構成条件を満たし、かつ、現在のクリーク中のすべてのテストと同時出題可能な、つまり共通項目数の少ないテストを、整数計画問題を利用し構成する。
4. 構成できた場合、これを C に追加する。このときの C を L へ追加する。与えられた計算時間を過ぎていない場合 3.へ
5. C 中のテストをランダムにいくつか削除する。このときの C を L へ追加する。与えられた計算時間を過ぎていない場合 3.へ
6. L 中から条件を最も良く満たす複数等質テストを探索し出力する

また、この探索で使用する整数計画問題は図4である。

入力

現在探索中のクリーク C 、項目重複上限数、テスト項目数、テスト構成条件 (テスト情報量上限 $UB(\theta_l)$ 下限 $LB(\theta_l)$)

変数

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

maximize

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

subject to

$$\sum_{i=1}^n x_i y_{k,i} \leq (\text{項目重複上限数})$$

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

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

$$\sum_{i=1}^n I_i(\theta_l) x_i = I(\theta_l)$$

$$LB(\theta_l) \leq I(\theta_l) \leq UB(\theta_l)$$

$$(l = 1, \dots, L)$$

ただし、定数 $y_{k,i}$ は以下のとおりである。

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

* ここで $\lambda_i (i = 1, 2, \dots, n)$ は互いに独立な $[0, 1)$ の連続一様分布からの乱数であり、本問題が解かれるたびにリサンプリングされるものとする

図4 隣接頂点探索用整数計画問題

本アルゴリズムでは、探索中に見つかったクリークに加えてもクリークとなるような頂点を、整数計画問題を利用し探索し、クリークを拡張していく。また、そのような頂点が見つからない場合(つまり、現在見つかったクリークが極大クリークである場合)そのクリーク中からランダムに頂点を削除し、再び、加えてもクリークとなる頂点を探索し、クリークを列挙していくアルゴリズムとなっている。つまり、現在あるクリークと似た(いくつかの頂点が追加削除された)クリークを探索していく局所探索アルゴリズムとなっている。

4. 研究成果

まず従来手法[3-6]と露出について比較した。表1は実アイテムバンクとシミュレーションアイテムバンク3個との構成テスト数・最大露出項目数、露出率(最大露出項目数/構成テスト数)を比較した物である。

表中のBSTは従来手法[4]、GAは[5]、BAは

[6]、RndMCPは[3]の手法を表している。また表中の $|C|$ はテスト構成数を、 Ex_C は最も露出が多かった項目の露出回数を、 $|C|/Ex_C$ は露出率を示している。露出率は最大でどれだけの割合のテスト数におなじ項目が出題されているかを表す指標である。また表中の OC は許されるテスト間の最大共通項目数を表している。それぞれのアイテムバンクから25項目のテストを構成する条件でこれらの比較を行った。

表4を確認すると、提案アルゴリズムはこれまで最も多くテスト構成できていた研究代表者らの手法[3]よりも多くのテストを構成でき、かつ、項目露出率も減っている結果となっている。さらに手法[3]では従来手法[4]よりも構成数が少ないテスト構成条件もあったが、本提案アルゴリズムではそのような条件でも最も多くのテストを構成できている。このような条件はテスト構成条件が厳しい条件で有り、ランダムにテストを構成していくよりも、整数計画問題を用いて探索を行う方が、条件が厳しい場合には構成テスト数が多くなったことが原因と考えられる。

一方で、テスト構成条件の緩いアイテムバンクサイズが大きな時や、 OC 値が大きな場合は、従来手法[3]とテスト構成数、項目露出率に大きな変化はなかった。今回与えた計算時間が大規模なテスト構成に対して短く、露出項目数を少なくする側の探索があまり行われなかったことが原因と考えられる。今後の課題としてはより探索効率の良いアルゴリズムを開発し、露出を減らすことと、テスト構成数を増加することを同時に可能とするようなアルゴリズムを開発することと考えられる。

本研究成果は分野のトップカンファレンスであるAIEDに採録された(学会発表、)。またこの研究アルゴリズムを、テスト構成数を多くする目的で用いた結果は電子情報通信学会の論文誌に掲載され(雑誌論文)2018年6月に電子情報通信学会論文賞を受賞予定である。

また本研究の成果は社会的にも高いインパクトを有しており、特に入試などの分野に実用化が強く期待されている。2020年の入試改革によりアイテムバンク方式でのeテストの導入が検討されており、既にリクルート社のSPIや医学系大学共用試験などにおけるCBT等に本研究を適用する実践も開始しており、本研究成果はますます波及すると考えられる。

<参考文献>

- [1] 石井 隆稔, ソンムアン ポクポン, 植野 真臣, 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, J97-D, pp.270-280.
- [2] T. Ishii, P. Songmuang, and M. Ueno, "Maximum clique algorithm for uniform test forms," The 16th International

表1 各手法とのテスト構成数、露出、露出比率の比較

Pool Size	OC	BST			GA			BA			RndMCP			Proposal		
		C	Ex _C	$\frac{EXC}{ C }$	C	Ex _C	$\frac{EXC}{ C }$	C	Ex _C	$\frac{EXC}{ C }$	C	Ex _C	$\frac{EXC}{ C }$	C	Ex _C	$\frac{EXC}{ C }$
500	0	12	1	8.3%	3	1	33.3%	5	1	20.0%	10	1	10.0%	17	1	5.9%
	5	20	6	30.0%	23	5	21.7%	96	16	16.7%	4380	362	8.3%	10736	823	7.7%
	10	20	12	60.0%	21	7	33.3%	107	15	14.0%	99983	12995	13.0%	100975	13141	13.0%
1000	0	21	1	4.8%	4	2	50.0%	6	2	33.3%	17	1	5.9%	33	1	3.0%
	5	40	14	35.0%	17	5	29.4%	104	11	10.6%	46305	3399	7.3%	48055	3504	7.3%
	10	40	21	52.5%	19	4	21.1%	105	14	13.3%	100000	8705	8.7%	101000	8630	8.5%
2000	0	53	1	1.9%	8	1	12.5%	12	1	8.3%	32	1	3.1%	70	1	1.4%
	5	80	27	33.8%	22	4	18.2%	104	6	5.8%	96876	3935	4.1%	97826	3829	3.9%
	10	80	43	53.8%	23	4	17.4%	103	7	6.8%	100000	4013	4.0%	100957	4014	4.0%
978 (actual)	0	24	1	4.2%	9	1	11.1%	9	1	11.1%	16	1	6.3%	31	1	3.2%
	5	39	10	25.6%	283	23	8.1%	371	23	6.2%	40814	2177	5.3%	44105	2163	4.9%
	10	39	13	33.3%	286	22	7.7%	381	24	6.3%	100000	5598	5.6%	101000	5274	5.2%

Conference on Artificial Intelligence in Education, pp. 451–462, 2013.

[3] T. Ishii, P. Songmuang and M. Ueno, "Maximum Clique Algorithm and Its Approximation for Uniform Test Form Assembly," in IEEE Transactions on Learning Technologies, vol. 7, no. 1, pp. 83-95, 2014.

[4] W.J. van der Linden, Liner Models for Optimal Test Design, Springer, 2005.

[5] 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.2, no.21, pp.141-161, 2008.

[6] P. Songmuang and M. Ueno, "Bees algorithm for construction of multiple test forms in e-testing," IEEE Transactions on Learning Technologies, vol.4, pp.209-221, 2011.

5. 主な発表論文等
(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 2 件)

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”, 電子情報通信学会論文誌 D Vol.J100-D(1), pp.47-59, 2017.[査読有]

石井隆稔, 植野真臣, “e テスティングにおける複数等質テスト自動構成手法の展望”, 日本テスト学会誌 Vol.11(1), pp.131-149, 2015. [査読有]

[学会発表](計 4 件)

Takatoshi Ishii, Maomi Ueno, “Algorithm for Uniform Test Assembly Using a

Maximum Clique Problem and Integer Programming”, Artificial Intelligence in Education 2017, 2017.(国際会議)

石井隆稔, 赤倉 貴子, 植野 真臣, “複数等質テスト構成のための整数計画問題を用いた乱択近似法の改善”, 日本教育工学会第 32 回全国大会, 2016.

石井隆稔, 赤倉貴子, 植野真臣, “最大クリーク問題と線形計画問題を用いた複数等質テスト自動構成手法”, 第 15 回情報科学技術フォーラム, 2016.

Takatoshi Ishii, Maomi Ueno, “Clique Algorithm to Minimize Item Exposure for Uniform Test Forms Assembly”, Artificial Intelligence in Education 2015, 2015.(国際会議)

6. 研究組織
(1)研究代表者
石井 隆稔 (ISHII, Takatoshi)
東京理科大学・工学部情報工学科・助教
研究者番号: 40741591