

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 8 日現在

機関番号：12612

研究種目：挑戦的萌芽研究

研究期間：2012～2014

課題番号：24650549

研究課題名(和文) e テスティングにおける複数等質テストの自動生成システムの開発

研究課題名(英文) Uniform Test Form Assembly System in e-Testing

研究代表者

植野 真臣 (Ueno, Maomi)

電気通信大学・その他の研究科・教授

研究者番号：50262316

交付決定額(研究期間全体)：(直接経費) 2,900,000円

研究成果の概要(和文)：本科学研究ではeテスティングにおける複数等質テスト自動構成手法を提案・開発した。複数等質テストとは、それぞれのテストに含まれるテスト項目は異なるが、統計的な性質が等しいテスト群である。本手法の特徴は、複数等質テスト構成を最大クレーク問題として解くことで、与えられたアイテムバンク・テスト構成条件で最大数のテストを構成可能である。また、限られた計算量でテスト構成を行う乱数探索を用いた近似手法を提案した。これにより、厳密法の指数時間計算量と多項式空間計算量を定数オーダーへと軽減できた。シミュレーション及び実データを用いた実験を行い、他手法より多くのテストを構成できることを示した。

研究成果の概要(英文)：Educational assessments occasionally require uniform test forms for which each test form comprises a different set of items, but the forms meet equivalent test specifications (i.e., qualities indicated by test information functions based on item response theory). We propose two maximum clique algorithms for uniform test form assembly. The proposed methods can assemble uniform test forms with allowance of overlapping items among uniform test forms. First, we propose an exact method that maximizes the number of uniform test forms from an item pool. However, the exact method presents computational cost problems. To relax those problems, we propose an approximate method that maximizes the number of uniform test forms asymptotically. Accordingly, the proposed methods can use the item pool more efficiently than traditional methods can. We demonstrate the efficiency of the proposed methods using simulated and actual data.

研究分野：教育学

キーワード：eテスティング

1. 研究開始当初の背景

近年、情報技術の普及により、様々な試験が情報技術を用いた評価により実施されている。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・テストにおいて情報技術を用いた評価の導入・運用が始まっている。ハイステークス・テストとは、結果が受験者に大きな影響を及ぼすテストを指す。このようなテストでは、異なる項目により構成されているにもかかわらず、各テストが等質であるようなテストが必要となる。例えば、毎年行われる資格試験では毎回の難易度が大きく異なってはならない。既に現在、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 (2007)(対応国内規格 JIS X7221:2011,2011) で、ハイステークス・テストでの複数等質テストが条件として記載されている。日本国内においては ISO/IEC と JIS 認定は未だにないが、情報処理技術者試験 IT パスポートは ISO/IEC 規格を参考に作成されており、条件を満たしている。ハイステークス・テストでの複数等質テストの重要性は今後も増し続けると予想できる。しかし、複数等質テスト構成の計算量は NP 困難であり、日本で最も多いアイテムバンクのサイズである 2万 3 万項目数から数百のテストを構成することが限界であった。

2. 研究の目的

しかし、近年、さまざまに分散した場所で、いつでも受験できる環境を提供することを想定するとほぼ 1 年で異なるテスト実施が尽きてしまうという問題がある。本研究課題では、コンピュータサイエンスの最新技術を用い、厳密解を求めることによって、この限界を大きく改善する。具体的には、等質テストの構成が最大クリーク問題に帰着することを提案し、厳密に最大数の等質テストを生成する手法を提案し、その有効性を示した。また、この手法の計算コストの問題点を近似アルゴリズムを用いることで解決した。

3. 研究の方法

本科研での提案手法はテスト間に項目の重複を許した条件で、最大数のテストを構成する手法である。

具体的には以下のような最大クリーク問題と呼ばれる最適化問題として複数等質テスト構成を行う。

変数

$$C \subseteq V \tag{1}$$

目的関数

$$\text{Maximize } |C| \tag{2}$$

制約関数

$$\forall v \in V, \tag{3}$$

(テスト v は与えられたテスト条件を満たす)

$$v \neq w, \forall v, w \in C, \{v \cap w\} \in E \tag{4}$$

ここで、 V は与えられたアイテムバンクとテスト構成条件から構成可能な全てのテストである。また E は項目重複条件を満たす 2 つのテストの組を格納しており ($\{v, w\} \in E \Leftrightarrow |v \cap w| \leq \text{overlap}$)、式(4)は、テスト集合 C はその中のどの 2 つのテストの組もその E に入っている、つまり、項目重複条件を満たすことを表す制約である。従って、この最適化問題は、どの二つも重複条件を満たすテスト集合において、含まれるテスト数が最大であるものを探索する。

この最適化問題はあるグラフからの最大クリーク問題となる。この最適化問題を表す模式図が図 1 である。

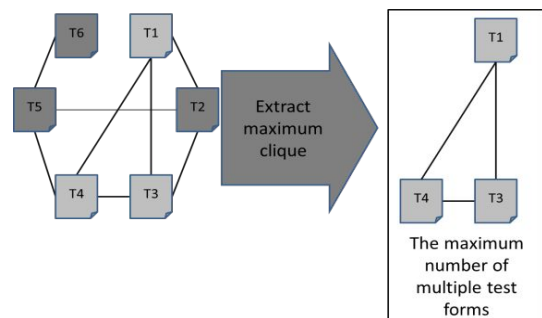


図 1 テスト構成のため最大クリーク問題

図 1 の頂点 ($\{T1, T2, T3, T4, T5, T6\}$) はテスト構成条件を満たす全てのテストである。また、図 1 中の二つのテストが重複条件を満たす場合、その間に辺をもつ。例えば、 $T1$ と $T2$ は互いに等質で重複条件を満たすため、その間に辺をもつ。式(1) ~ (4)の最適化問題中の V と E はこのグラフの頂点と辺に対応している。つまりこの最適化問題はこのようなグラフ構造(例えば、図 1 中左の構造)中のどの二つの頂点も辺で結ばれている構造(クリーク、完全グラフ構造と呼ばれる)のうち最も要素数が大きなもの(図 1 中の右の構造)を探索している。

本提案手法は、テスト情報量の上限下限の制約を用いて等質性を取り扱う。石井・ソムアン・植野(2014)は構成テスト間に項目重複を許した場合でも最大数の複数等質テストを構成でき、項目重複を許した場合の手法と比較しても、数十倍以上のテストを構成できる。ただし石井・ソムアン・植野(2014)には計算コストが非常に大きい問題がある。この問題を、近似アルゴリズムを用いることで解決した手法が Ishii, Songmuang and Ueno (2014) である。この手法は乱択アルゴリズムを用いることで、時間漸近的に最大数のテストを構成可能である。また、近似アルゴリズムを用いるため、任意の時間でテスト構成が可能である。具体的には、アルゴリズム 1 を用いてテスト構成を行う。

アルゴリズム 1

Ishii, Songmuang and Ueno (2014)

Input
 アイテムバンク, テスト構成条件
 計算量に関するパラメータ: C1, C2, C3

Output
 複数等質テスト

Step 1:
 与えられたテスト構成条件の重複条件以外を用い, それらを満たすテスト可能テストを C1個構成する.

Step 2:
 Step 1で構成した C1個のテストをグラフの頂点とする. またこのグラフ中のすべての 2頂点(すべてのテストのペア)について重複項目数を数え, それが与えられた重複項目条件以下ならその頂点間に辺を引く.

Step 3:
 Step 2で構成したグラフより最大クリークを抽出する. ここでは最大クリークの探索を与えられた計算時間 C2だけ行う. 得られた最大クリークと現状見つけている最大クリークとを比較し, より大きなものを現状見つけている最大クリークとして保存しておく.

Step 4:
 計算開始からの経過時間が C3を上回っていれば計算を終了し, 現状見つけている最大クリークを出力する.
 それ以外の場合, Step 1へ戻る.

Ishii, Songmuang and Ueno (2014)のアルゴリズムは任意の計算コスト(C1 と C3)で計算が可能である. これは, 石井・ソムアン・植野(2014)の時間的計算量 $O(2^{n0.19171F})$ と空間的計算量 $O(F^2)$ に対して格段に扱いやすい.

4. 研究成果
 従来手法である進化的アルゴリズムを用いた同時構成手法(例えば, Sun, Chen, Tsai and Cheng, 2008)と Big Shadow Test 法(van der Linden, 2005)は, いずれも同時構成法を近似化した手法である. 進化的アルゴリズムを用いた手法は近似解法を用いることで, また Big Shadow Test 法は最適化問題を簡易化することで計算量を軽減している. 進化的アルゴリズムは任意の最適化問題を, 任意の計算時間で計算できる. そのため, これらのアルゴリズムを用いてテスト構成を行うことによって, 任意の計算時間でテスト構成が可能となる. また, 例えば, 項目識別力と平均難易度の両方を同時に最適化するような複雑な最適化を行う新しいテスト構成手法も提案されている(例えば, Ho, Yin, Hwang, Shyu, and Yean, 2009). ただし, 同時構成手法はどれもテスト構成数を最大化できない問題がある. Big Shadow Test 法は LP を用いた同時構成法の最適化問題を一部簡略化することにより, 比較的低い計算コストで複数等質テストを構成できる手法であり, 時間的計算コストは $O(2^{2n})$ である. また LP を用いてテスト構成を行う手法であるため, 容易に実装が可能である. このような点から, 大規模なアイテム

バンクから, 決められた数のテストを構成する場合, 最も有効な手法の一つであると考える.

Songmuang and Ueno(2011)の手法は, できる限り多くのテストを構成できる最初の実用的な手法である. この手法の特徴は, 等質性とテスト構成数を同時に含んだ目的関数を最適化することにより, 等質で多くのテストを構成可能である点である. この手法は並列計算を用いることで実質的な計算量を軽減できる.

表1 各手法のテスト構成数の比較

(Ishii, Songmuang and Ueno, 2014)

Item Bank Size	OC	Methods			
		BST	GA	BA	RM
500	0	12	3	5	10
	5	20	23	96	4380
	10	20	21	107	99983
1000	0	21	4	6	17
	5	40	17	104	46305
	10	40	19	105	100000
2000	0	53	8	12	32
	5	80	22	104	96876
	10	80	23	103	100000
978 (actual)	0	24	9	9	16
	5	39	283	371	40814
	10	39	286	381	100000

表1は, Ishii, Songmuang, and Ueno (2014)で示されている同一のアイテムバンク, 同一条件下での各種法のテスト構成数を示したものである. 表3は与えられたアイテムバンク(500項目, 1000項目, 2000項目のシミュレーションデータと978項目の実アイテムバンク)から25項目のテストを構成する条件で, 各手法を適用したときのテスト構成数を示している. それぞれの手法には6時間の計算時間を与えている. 表3中のItem Bank Sizeは与えたアイテムバンクの大きさを, OCは構成条件として与えたテスト間の最大項目重複数を表している. 比較されている手法は Big Shadow Test 法(van der Linden, 2005): BST, 遺伝的アルゴリズムを用いた手法(Sun, Chen, Tsai and Cheng, 2008): GA, Songmuang and Ueno (2011): BA, 最大クリーク問題を用いる近似手法(Ishii, Songmuang, and Ueno, 2014): RMである.

表3より, 提案手法が項目重複を許す条件(OC=5, 10)において他手法の数倍以上のテストを構成していることがわかる. また, Songmuang and Ueno(2011)の手法は, テスト構成数を含むパラメータを漸近的に最適化する手法であるため Big Shadow Test 法, 遺伝的アルゴリズムを用いた手法よりも多くのテストを構成可能である.

一方, 非重複条件(OC=0)においては, Big Shadow Test 法が最も多くテストを構成できている. つまり, 非重複条件でのテスト構成

では Big Shadow Test 法の近似化が、時間あたりの計算効率に優れることが示唆される。ただし、提案手法は漸近的に最大のテスト数を構成するため、さらに長い計算時間を与えると Big Shadow Test 法よりも多くのテストが構成可能となる。

また、重複条件 ($OC=5, 10$) においては Big Shadow Test 法、遺伝的アルゴリズムを用いた手法、Songmuang and Ueno(2011)は重複項目条件を緩和してもテスト構成数があまり増加していない。また、遺伝的アルゴリズムを用いた手法と Songmuang and Ueno(2011)は、シミュレーションアイテムバンクを用いた結果において、アイテムバンクサイズを大きくしてもテスト構成数があまり増加していない。アイテムバンクサイズの増加や重複条件の緩和は、最適化問題としての計算複雑性を増加させる。そのため、これらの手法は効率的な解探索を行うことが困難になったと考えられる。このような条件でも、Ishii, Songmuang and Ueno (2014)は多くのテストを構成できる。

ただし、Ishii, Songmuang, and Ueno (2014)でも、重複条件を緩和しても ($OC=5 \rightarrow 10$)、テスト構成数はさほど増加せず、10 万を超えていない。これは一度に探索を行う部分グラフの頂点数(計算コストパラメータ $C1$)に 10 万を設定したためである。すなわち、厳密な最大の重複等質テスト群は 10 万以上のテストを含んでいるが、計算コストの限界により 10 万までしかテストを構成できていない。従って、さらに大きな計算コストパラメータ $C1$ を設定することでよりテスト構成数を増加できると考えられる。

以上のような特徴がそれぞれの手法にはある。そのため、テスト管理者はこれらの手法を、状況に応じて使い分ける必要がある。概ね、以下のように使い分ける必要がある。(1)大規模なアイテムバンクから項目の重複を許し、大量のテストを構成する場合、Ishii, Songmuang, and Ueno (2014)の手法が構成テスト数と計算コストの点で最も実用的である。(2)ある程度潤沢なアイテムバンクからテスト間の項目重複を許さずテスト構成を行う場合は、計算コストや利便性、テスト構成数の点で van der Linden (2005)の Big Shadow Test 法が望ましい。本手法は汎用的な最適化ツールである IBM 社の CPLEX などを使用可能であり、実装が容易である。(3)非常に限定されたアイテムバンクから多くのテストが必要な場合は、Belov and Armstrong (2006)や石井・ソムアン・植野(2014)を、項目重複に対する制約により使い分ければよい。

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4件)

石井隆稔、ソムアン・ポクボン、植野真臣、(2014)、最大クリーク問題を用いた重複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, (2), 270-280, 査読有。

Ishii, T., Songmuang, P., & Ueno, M., (2014), Maximum Clique Algorithm and Its Approximation for Uniform Test Form Assembly, IEEE Transactions on Learning Technologies, 7, (1), 83-95. 査読有。

[学会発表](計9件)

Takatoshi Ishii, Pokpong Songmuang, Maomi Ueno., Maximum Clique Algorithm for Uniform Test Forms Assembly, AIED(Artificial Intelligence in Education)2013, 2013年07月09日~2013年07月13日, Memphis(United States).

石井隆稔, ソムアン ポクボン, 植野真臣., テスト構成数を最大化する重複等質テスト自動構成手法, 日本教育工学会, 2012年09月15日~2012年09月17日, 長崎大学(長崎県)

6. 研究組織

(1)研究代表者

植野 真臣 (UENO, Maomi)
電気通信大学・大学院情報システム学研究科・教授
研究者番号: 50262316

(2)研究分担者

()

研究者番号:

(3)連携研究者

森本 康彦 (MORIMOTO, Yasuhiko)
研究者番号: 10387532

橋本 貴充 (HASHIMOTO, Takamitsu)
研究者番号: 10387532

ソムアン ポクボン (SONGMUANG, Pokpong)
研究者番号: 50613874

安藤 雅洋 (ANDO, Masahiro)
研究者番号: 00345539