K-005

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
Uniform Test Assembly using Maximum Clique Algorithm and Linear Programming

石井 隆稔 † 赤倉 貴子 † 植野 真臣 ‡
Takatoshi Ishii Takako Akakura Maomi Ueno

1. はじめに

2007 年には国際標準規格 ISO/IEC 23988:2007[1] (国内規格 JIS X7221:2011[1])が策定され、今後 e テスティングの普及が見込まれる。この規格の中で、その結果が受験者に大きな影響を与えるテストでは "複数等質テスト" を用いることが推奨されている。複数等質テストとは各テストに含まれる項目は異なるが、推定得点の予測誤差が等しいテスト群である。これは同じ問題が出題されることによる問題流出リスクを避けかつ、それぞれのテストの難易度や能力推定精度を一定に保つためである。これまで複数等質テストはテスト管理者・作問者の勘と経験により構成されてきたが、近年計算機技術の発達により複数等質テストの自動構成が可能となりつつある[1]。

一般に e テスティングではテストの管理手法として、出題可能な問題（以降項目と呼ぶ）データベース：アイテムバンクを用いる手法が用いられる。アイテムバンクには項目の内容やの能力推定精度などが格納され、これらの情報を使って構成されたテストの能力測定精度や得点分布などを推定可能である。テストの自動構成は、これらの情報を利用し所望の性質を持つ項目の組み合わせ（つまり、テスト）を組み合わせ最適化問題により探索することである。このような複数等質テスト構成手法の研究で重要な課題は、可能な限りテスト間の項目の重複を抑え、可能な限りテストの構成数を増やすことである。。

この目的のため、著者らは先行研究[3]において、テスト構成を、グラフ上で定義される最適化問題である最大クリ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・項目重複条件で従来手法の数十から数百倍のテストを構成可能な手法を提案した。しかし、本手法は必要となる空間計算量が多く、構成テスト数が計算機環境により制限されやすい問題がある。

本研究ではこの問題を緩和し、同一の計算機環境、アイテムバンク、テスト構成条件でより多くのテストを構成可能な手法を提案する。具体的には、テスト構成のための最大クリーク探索中で、対象のグラフ構造を探索の必要に応じて構成することにより、グラフ構造の保持に必要であった空間計算量を削減する。最後にこの手法の有効性をシミュレーションデータ及び実データを用いた実験により示す。

2. 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

2.1 複数等質テスト構成のための最大クリーク問題

本手法は複数等質テスト構成を最大クリーク問題 (Maximum clique problem) として解く。具体的には、 以下のグラフ中から最大クリークの探索・抽出を行い、 複数等質テストを構成する。

a. **頂点：** 与えられたアイテムバンクから構成可能な, テストの構成条件を満たすテスト(以降,”可能テスト" と呼ぶ) 全てを頂点とする。

b. **辺：** 二つの可能テストが項目重複条件を満たす (項目重複数が項目重複条件以下である) 場合その二つの頂点(テスト) 間に辺を引く。

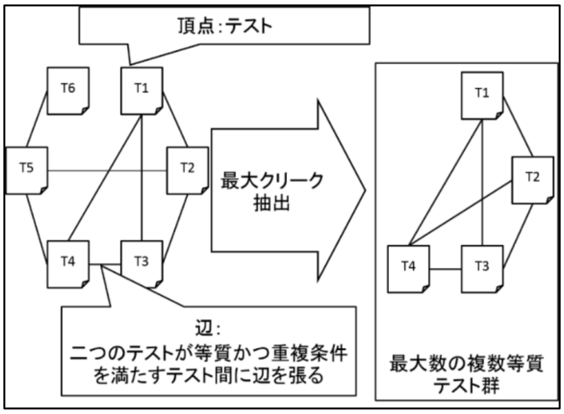

図 1 複数等質テスト構成のためのグラフ構造

図 1 はこのようにして構成したグラフ構造の例である。このグラフ中のクリークは複数等質テストである。従って、このグラフ中の頂点数が最大のクリークは構成可能な最大の複数等質テスト群となる。ただし、このように構成したグラフは頂点数が非常に大きくなり、厳密に最大クリークを探索することは困難である。そこで本研究では近似的により大きなクリークを探索するアルゴリズムを提案する。

2.2 アルゴリズム

具体的には、以下のようなアルゴリズムを用いてグラフ中の最大クリークを探索する。

STEP 1. 初期化
先行研究[3]を用いて初期解の探索を行う。発見した複数等質テストを暫定解とする。

STEP 2. 頂点追加
後述する線形計画問題（より詳しくは整数計画問題）を用いて暫定解中のすべてのテストと等質なテストを構成する。この線形計画問題に解があればその解を、暫定解へ追加する。また、追加した個数をカウントする。

STEP 3. 頂点の削除
暫定解から、テストを STEP2 で追加した個数の 10%削除する。計算開始からの計算時間が与えられた計算時間以内であれば STEP2 へ戻る。

STEP2 で用いられる線形計画問題（より詳しくは整数計画問題）は以下の通りである。

† 東京理科大学
‡ 電気通信大学

変数

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

目的関数

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

制約

$$\mathrm{I}(\theta) = \sum_{i=1}^{n} \mathrm{x_i I}_i(\theta)$$
$$\mathrm{LB}(\theta_l) < \mathrm{I}(\theta_l) < \mathrm{UB}(\theta_l) \tag{2}$$
$$(\mathrm{l} = \{1, \ldots, \mathrm{L}\})$$

$$\sum_{i=1}^{n} x_i = (\text{所望のテスト項目数}) \tag{3}$$

$$\sum_{i=1}^{n} x_i \mathrm{x}'_{ji} \leq (\text{項目重複数上限})$$

ただし、

$$x'_{ji} = \begin{cases} 1 & i\text{番目の項目が暫定解中のテスト}j\text{に含まれる} \\ 0 & \text{それ以外} \end{cases} \tag{4}$$

$$(\mathrm{j} = \{1, \ldots, (\text{暫定解中のテスト数})\})$$

ただし、$\lambda_i \in [0,1)$は互いに独立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。また、

$$\mathrm{I_i}(\theta) = a_i^2 \mathrm{P_i}(\theta)(1 - \mathrm{P_i}(\theta))$$
$$\mathrm{P_i}(\theta_j) = \frac{1}{1 + \exp(-\mathrm{D}a_i(\theta_j - b_i))}$$

$\mathrm{UB}(\theta_\mathrm{l}), \mathrm{LB}(\theta_\mathrm{l})$は$\theta_\mathrm{l}$での情報量関数の上限・下限である。

つまり、本線形計画問題は、式(2)：受検者の推定得点の予測誤差を評価するための項目反応理論に基づくテスト情報量I(θ)が$\mathrm{LB}(\theta) < \mathrm{I}(\theta) < \mathrm{LB}(\theta)$となり、式(3)：テスト項目数が等しく、式(4)：暫定解中のすべてのテストと項目重複数が与えられた項目重複上限以下の、式(1)：ランダムなテストを構成している。

先行研究[3]では、グラフ全域からランダムにC_1頂点を取り出した部分グラフからの探索を繰り返す手法であるため、空間計算量は$O(C_1^2)$である。またC_1以上の複数等質テストを構成できない。提案手法の空間計算量は内部で先行研究[3]を用いているため、$O(C_1^2)$であるが、提案手法はその後さらに構成テスト数を局所探索により改善する。そのため、同一の計算量でもより多くのテストを構成可能である。

3. 実験

最後に本手法の有効性を示すため、従来手法と提案手法でのテスト構成数を比較した。比較を行った手法は、BST 法[4], GA 法[5], BA 法[6], RndMCP 法[3] である。実データを模して発生させたシミュレーションアイテムバンク(アイテムバンクサイズ=500,1000,2000)と実アイテムバンク(アイテムバンクサイズ=978)から表 1 のようなテスト情報量への条件を持つテストを構成した。テスト中の項目数は 25 問、項目重複の上限は 0(重複項目なし)，5(テスト全体の 20%まで)，10(40%まで)としてテスト構成を行った。それぞれの構成手法には 24 時間の計算時間を与え、同一の計算機環境でテスト構成数を比較した。また RndMCP 法と提案手法には$C_1 = 100000$を与えた。これは使用した計算機環境が許す最大の空間計算量である

表 2　テスト情報量への条件

テスト情報量(上限/下限)				
$\theta = -2.0$	$\theta = -1.0$	$\theta = 0.0$	$\theta = 1.0$	$\theta = 2.0$
(2.0/2.4)	(3.2/3.6)	(3.2/3.6)	(3.2/3.6)	(2.0/2.4)

表 3　各手法のテスト構成数

アイテムバンクサイズ	重複条件	各手法のテスト構成数				
		BST	GA	BA	RndMCP	提案手法
500	0	12	3	5	10	17
	5	20	23	96	4380	14000
	10	20	21	107	99983	104525
1000	0	21	4	6	17	34
	5	40	17	104	46305	55837
	10	40	19	105	100000	105504
2000	0	53	8	12	32	70
	5	80	22	104	96876	101223
	10	80	23	103	100000	104604
978（実データ）	0	24	9	9	16	35
	5	39	283	371	40814	53080
	10	39	286	381	100000	104604

表 3 は各手法・条件でのテスト構成数を表している。提案手法は、特に重複条件=5 の場合、RndMCP と比較して大きくテスト構成数を増やしている。また重複条件=10 の場合やアイテムバンクサイズ 2000 での重複項目数=5 の場合、RndMCP は与えた計算量のコストC_1（10 万）を超えるテストを作ることができないが、提案手法は 10 万以上のテストを構成できている。この結果から提案手法は、同一の計算機環境でより多くのテストを構成可能と考える。

4. おわりに

本研究では複数等質テスト構成のための最大クリーク探索のための新しいアルゴリズムを提案した。提案手法は整数計画問題を用いてグラフを探索の必要に応じて構成することにより、従来手法と比較し空間計算量を減らした。これにより同一の計算機環境であれば、より広い部分グラフからの探索を可能とし、より多くのテストを構成可能である。今後の課題はさらに多くのテストを構成可能となるよう、効率的なアルゴリズムを開発したい。

謝辞

本研究の一部は科学研究費補助金(若手研究(B), 代表:石井隆稔)「e テスティングの信頼性向上の為の項目暴露を考慮した複数等質テスト構成手法の開発」(15K16260) および, (基盤研究(A), 代表:植野真臣)「大規模型 e テスティング・システムとその運営モデルの開発」(15H01772) の助成を受けた.

参考文献

[1] JIS, “JIS X7221:2011,”アセスメント提供における情報技術（IT）利用の規範”, (2011)

[2] 植野 真臣,永岡慶三, “e テスティング”, 培風館, (2009).

[3] T. Ishii, P. Songmuang, and M. Ueno, “Maximum clique algorithm and its approximation for uniformtest form assembly", IEEE Transactions on Learning Technologies, Vol.7, No.1, p.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).