

# e テスティングにおける複数テスト自動構成近似手法 Approximation of Multiple Test Forms Assembly using Maximum Clique Algorithm

石井 隆稔<sup>\*1</sup>, ソンムアン ポクボン<sup>\*2</sup>, 植野 真臣<sup>\*1</sup>  
Takatoshi ISHII<sup>\*1</sup>, Pokpong SONGMUANG<sup>\*2</sup>, Maomi UENO<sup>\*1</sup>

<sup>\*1</sup>電気通信大学 情報システム学研究科

<sup>\*1</sup>Graduate School of Information Systems, The University of Electro Communications

<sup>\*2</sup>タマサート大学 理工学部

<sup>\*2</sup>Faculty of Science and Technology, Thammasat University Pathum Thani, Thailand

Email: ishii@ai.is.uec.ac.jp

あらまし：本論文では、複数等質テストを自動構成する近似手法を提案する。複数等質テストとは、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、限られた計算コストの中で漸近的に最大数のテストを構成できる点である。具体的には、著者らの厳密手法を近似し、計算コストを改善した。また、シミュレーションデータを用いた実験を行い、本手法は他手法よりも統計的に多くのテストを構成できることを示した。

キーワード：e テスティング, 複数等質テスト自動構成, 最大クリーク問題

## 1. はじめに

実際のテスト構成では、しばしば複数の等質なテストが必要となる。例えば、資格試験などでは毎回の難易度が異なるようにテストの得点分布や所要時間が一定でなければならない。また、試験が選択式の部分テストで構成されている場合、それらはそれぞれ互いに等質でなければならない。

これまで等質なテストはテスト管理者の勘と経験により構成されてきた。しかし、近年、e テスティングの普及に伴い、アイテムバンク方式のテスト構成が一般化し、自動構成が可能となりつつある。

例えば、van der Linden は線形計画法を用いてテスト構成を行う Big Shadow Test (BST)法 [1]を提案した。この手法は、テストに出題する項目群とアイテムバンクに残す項目群を、線形計画問題を解き等質化すること逐次的に構成されるテスト群全体を間接的に等質化する。比較的低い計算量でテスト構成を行えたが、(1)構成テスト数が増えると後に構成したものと最初に構成したものの間の差異が大きくなる、(2)この手法で作られたテスト群が与えられた条件中で最大数のテスト群である保証がない、という二つの問題点があった。

BST の持つ前者の問題を解決するため、Sun らは遺伝的アルゴリズム(Genetic Algorithm)を利用する GA 法を提案した [2]。この手法は全てのテストを同時に構成することで、すべての構成テスト間の差異を最小化した。また、Songmuang らも Bees Algorithmを用い、GA法のパフォーマンスをさらに改善した [3]。

これらの手法(BST,GA,BA)はテスト間の差異は軽減したが、与えられたアイテムバンク・条件中で最大数のテストを構成することは不可能であった。つまり、どの手法もアイテムバンクを有効活用することは困難であった。

これを解決するために、Belov らは集合充填問題(Maximum Set Packing Problem)を用いて、与えられた条件・アイテムバンク中で最大数のテストを構成する手法(MSP 法)を提案した [4]。しかし、この手法は構成テスト間に項目重複を許さない条件でのみ使用可能であった。この条件は、言い換えれば全ての項目は最大で一度しか出題できない条件であるためテストの構成数を大きく制限しアイテムバンクの有効活用を阻害していた。

そこで著者らは、テスト間の重複を許す条件で与えられた条件・アイテムバンクから最大数のテストを構成する手法を提案した [5]。具体的にはテスト構成条件を満たすテスト(以下、可能テスト)を頂点とし、テスト間の重複関係を辺としたグラフからの最大クリーク問題としてテスト構成を定式化し、これを厳密に行う手法:ExMCP と漸近的に行う手法:RndMCP を著者らは提案した。

ExMCPはBelov [4]のMSP法をテスト間に項目重複を許した状態でもテスト構成可能に一般化した手法とも考えられる。Belov の手法と同様に出力されるテスト数はその構成条件中で最大数であることが数学的に保証できる。しかし、ExMCP の計算コストは可能テスト数に対し、指数的に増大し、現実的な規模でのテスト構成を行うことは困難であった。

RndMCP はExMCPをRandom Search Approachを用いて与えられたコスト制約で行えるように近似化した手法である。RndMCPの特徴はRandom Search Approachを用いることで漸的にテスト構成数を最大化する点である。出力されるテスト数が最大であることは保証できないが、多くの場合、RndMCPは先行研究(例えばBST, GA, BA, MSP等)よりも多くのテストを構成できる。また、ExMCPよりも計算コストが低く、ExMCPでは計算困難で条件でもRndMCPはテスト構成可能である。

表 1 テスト構成数の平均

Item Pool Size	OC	Constraint 1							
		BST		GA		BA		RM	
		Avg.	SD	Avg.	SD	Avg.	SD	Avg.	SD
80	0	7.07	1.60	7.96	1.73	8.93	1.60	11.15	2.49
	1	20.00	0.00	30.84	10.02	40.93	10.65	139.73	39.09
	2	20.00	0.00	52.63	17.97	76.48	18.25	2446.21	800.15
100	0	8.57	1.55	9.57	1.87	10.89	1.81	14.12	2.47
	1	25.00	0.00	35.00	10.96	55.04	12.28	218.32	52.40
	2	25.00	0.00	54.27	16.82	91.31	21.80	4696.10	1256.01
120	0	10.60	1.91	11.05	2.00	13.38	1.90	17.48	3.09
	1	30.00	0.00	40.41	11.67	70.94	15.88	318.26	66.74
	2	30.00	0.00	55.11	16.63	110.20	24.98	7933.48	1857.89

表 1 比較手法より多くのテストを構成した回数

Item Pool Size	OC	Constraint 1								
		vsBST			vsGA			vsBA		
		>	=	<	>	=	<	>	=	<
80	0	0	0	100	0	1	99	0	9	91
	1	0	0	100	0	0	100	0	0	100
	2	0	0	100	0	0	100	0	0	100
100	0	0	0	100	0	0	100	0	2	98
	1	0	0	100	0	0	100	0	0	100
	2	0	0	100	0	0	100	0	0	100
120	0	0	0	100	0	0	100	0	1	99
	1	0	0	100	0	0	100	0	0	100
	2	0	0	100	0	0	100	0	0	100

前回の発表では、実データである7アイテムバンクに対し、それぞれの手法でのテスト構成数を比較した。しかし、実験に使用したアイテムバンク数が少なく、提案手法のパフォーマンスを統計的に、つまり、アイテムバンクに対し一般化した評価を行えていなかった。そこで本稿では、近似提案手法:RndMCPの有効性を統計的に示すため、シミュレーションにより発生させた十分多数(それぞれの条件で100 ずつ)のアイテムバンクに対し、テスト構成を行いすべての条件中で提案手法は主要な先行研究手法よりも多くのテストを構成することを確認した。

## 2. 提案手法

著者らのテスト構成手法はテスト構成をグラフ論のよく知られた組み合わせ最適化問題である最大クリーク問題として行う。以下のようなグラフ中から最大クリークの探索・抽出を行うことによって、テスト構成を行う。

- 頂点：重複条件を除く全てのテスト構成条件を満たす、与えられたアイテムバンクから構成可能なすべてのテスト(以後、可能テスト群と呼ぶ)
- エッジ：二つの可能テストが重複条件を満たす場合(重複条件により指示される最大重複項目数より重複項目が少ない場合)その二つの頂点(テスト)間にエッジを張る

これを厳密に行う手続きが ExMCP である。ExMCP は与えられたアイテムバンクから最大の複数等質テストを構成することが数学的に保証できるが、その時間的・空間的計算量はそれぞれ  $O(2^F)$ ,  $O(F^2)$  となる。ただし  $F$  は可能テスト数である。

RndMCP は ExMCP を Random Search Approach を用いて近似化した手法である。RndMCP では計算コストによる条件が許す個数だけ可能テスト群をランダムサンプリングし、その中から最大クリーク探索を時間の許す限り繰り返す。これにより、与えられた構成条件・アイテムバンク中から時間漸近的に最大数の複数等質テスト群を出力する。

## 3. 評価実験

本手法の有効性を示すため実験を行った。シミュレーシ

ョンにより、80,100,120 のサイズのアイテムバンクそれぞれ 100 個発生させ RndMCP と従来手法のテスト構成数を比較した。

表 1 からは RndMCP のテスト構成数の平均はその他の手法よりも多いことがわかる。また、各試行での RndMCP がその他の手法より多くのテストを構成した回数をまとめた表 2 比較手法より多くのテストを構成した回数の結果は、全ての場合において RndMCP が他の手法に比べ等しいか、それより多くのテスト構成したことがわかる。

## 4. おわりに

本稿では複数等質テスト自動構成手法の統計的な性質について評価を行った。実験によりさまざまな場合において主要な先行研究手法に比べ、多くのテストを構成でき、アイテムバンクを有効活用できることを示した。

## 参考文献

- [1] Wim J. van der Linden., Linear Models for Optimal Test Design, 2005.
- [2] Koun-Tem Sun , Yu-Jen Chen , Shu-Yen Tsai and Chien-Fen Cheng., "Creating IRT-Based Parallel Test Forms Using the Genetic Algorithm Method," 2008.
- [3] Pokpong Songmuang and Maomi Ueno., "Bees Algorithm for Construction of Multiple Test Forms in E-Testing," 2011.
- [4] Dmitry I. Belov and Ronald D. Armstrong, "A Constraint Programming Approach to Extract the Maximum Number of Non-Overlapping Test Forms," 2006.
- [5] 石井隆稔, ソムアンボクボン, 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," 第37回 教育システム情報学会 全国大会, 2012.
- [6] T. Ishii, P. Songmuang , M. Ueno, "Maximum Clique Algorithm for Uniform Test Forms Assembly," The 16th International Conference on Artificial Intelligence in Education, 2013.