

項目露出率を最小化する複数等質テスト構成手法

石井 隆稔*, 舩田 博之**, 仁田 光彦**, 植野 真臣*

電気通信大学大学院 情報システム学研究所*, 株式会社リクルートキャリア**

1 はじめに

実際のテスト構成では、しばしば複数等質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らないように、テストの統計的な性質、例えば、得点分布や所要時間が一定でなければならない。これまで複数等質テストはテスト管理者の勤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、テストの自動構成が可能となりつつある。

一般に、e テスティングでは、テストの管理手法としてアイテムバンク方式が用いられる。これは、コンピュータで出題する問題（以降、項目と呼ぶ）を管理するデータベースであり、項目ごとの出題分野や統計データなどを格納している。テストの自動構成とは、これらのデータを利用して、所望の性質を持つよう項目の組み合わせを計算機により最適化する。

例えば、Belov と Armstrong はテスト構成を集合充填問題 (Maximum Set Packing) 問題として解いている [1]。彼らはテストをそれぞれ排他的な等質条件を満たす項目の集合と定義し、アイテムバンクを最大数のテストへ分割する問題として解いている。そのため、与えられたアイテムバンク・テスト構成条件および構成テスト間に項目重複を許さない条件において、数学的に証明可能な最大数のテスト群を構成可能である。しかし、この手法は計算コストが高く実用的なサイズのテスト構成を行うことは困難であった。

そこで Belov はこの手法をより扱いやすい形で近似した手法を提案した [2]。この手法は線形計画法 (Linear Programming) を用いたテスト構成を繰り返すことで、重複のない等質なテスト群を構成した。しかし、これまで紹介した手法は、構成テスト間に項目の重複を許しておらずそれぞれの項目は最大でも一度ずつしか出題できない。そのため、テストの構成数が大きく制限される問題があった。つまり、この条件はアイテムバンクの有効活用を阻害していた。

そこで、著者らは構成テスト間に項目重複を許した条件で、与えられたアイテムバンク・構成条件中で最大数のテストを構成し、アイテムバンクを有効活用する手法を提案した [3]。この手法は Belov と Armstrong の手法 [1] を、項目の重複条件について一般化し、テスト間の重複関係をグラフとしてとらえた最大クリーク問題としてテスト構成を行う。これによ

り、この手法は複数テスト自動構成手法の中で最も多くの等質テストを構成でき、アイテムバンクをより有効活用できた。しかし、この手法の構成テスト群には、項目の露出 (exposure, 出題回数) に偏りが生じ得る。例えば、ある問題は、構成されたテストの 15% に出題されているが、別の問題は構成されたテストのたった 2% にしか出題されない状況が生じ得る。（この“同じ項目が構成テストのどの程度の割合に出題されているか”を本論文では、以降、露出率:Item Exposure Rate と呼ぶ。）

一般にテスト運用上、このような項目使用回数・露出率の偏りは好ましくない。なぜなら、露出の多い項目は受験者間で共有されやすく、経年による運用において、その項目の信頼性が失われやすくなるためである [4]。適応型テストの研究では、これを回避するための様々な手法が提案されている (ex, [5])。

そこで本研究では、複数等質テスト構成において露出率を制御可能な手法を提案する。具体的には、著者らの手法 [3] へ項目露出率に関する条件を追加した一種のクリーク問題として解くことで、最大項目露出率が最小な複数等質テスト群を構成する。特徴は、漸近的に項目の最大露出率を最小化できる点であり、多くの場合、以前の手法 [3] よりも少ない露出率で同等のテスト数を構成可能である。本稿ではこの手法の有効性を実データを用いた実験で示す。

2 提案手法

本研究ではテスト間に重複を許した条件で最大項目露出率が最小となる複数等質テスト構成手法を提案する。具体的には、著者らの以前の手法 [3] と同様に、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、一種のクリーク問題として行う。

提案手法は、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群（以降、可能テストと呼ぶ）、辺を重複条件の満足としたグラフから、露出率最小のクリーク探索問題として定式化する。

図 1 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。このようなグラフ中においてクリークは等質で重複条件を満たすテスト群を表す。

例えば、サイズが 3 以上のクリーク $\{ \{ T1, T2, T3,$

表 1: テスト構成数と項目の最大出題回数, 露出率の関係

重複条件	提案手法			以前の手法 [3]		
	構成数	出題回数	露出率	構成数	出題回数	露出率
2122111	235	27	11.49%	238	31	13.03%
1222111	231	26	11.26%	233	31	13.30%
1122211	328	37	11.28%	329	38	11.55%
1122121	243	34	13.99%	247	38	15.28%
1122112	274	32	11.68%	279	35	12.54%

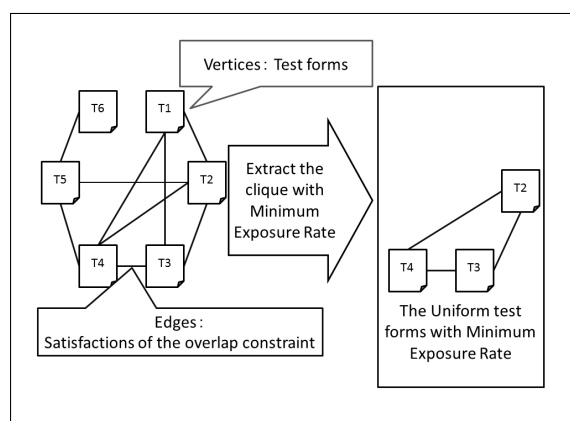


図 1: 本提案手法の模式図

$\{T4\}$, $\{T1, T2, T3\}$, $\{T1, T2, T4\}$, $\{T1, T3, T4\}$, $\{T2, T3, T4\}$, $\{T2, T4, T5\}$ はそれぞれ等質で重複条件を満たすテスト群である。以前の手法 [3] はこのグラフ中から最大クリーク抽出し, 最大数の複数等質テスト群を探索した。提案手法ではこれらのクリークの中から最小の項目露出率を持つテスト群を探索しテスト構成を行う。ただし, 厳密な計算には莫大なコストが必要なため, Random Search Approach や計算の打ち切りを行い, 時間漸近的に最小露出率を持つクリークを探索する。

この探索には, Nakanishi and Tomita のアルゴリズム [6] に露出率に関する条件を加える形で改変した新しいアルゴリズムを用いている。(本稿ではスペースの都合上省略する)

3 実験

本手法の有効性を示すため実験を行った。実験に使用したアイテムバンク・テスト構成条件は, 主に企業の採用選考場面で使用される株式会社リクルートキャリアが提供する WEB テストのものであり, アイテムバンクは 7 分野, 合計で約 1000 の項目を持つ。テスト構成条件は項目反応理論によるテスト情報量の上限・下限, また, それぞれの分野内での最大重複

項目数が設定されている。これらの条件の下, 提案手法と以前の手法 [3] でテスト構成数, 項目出題回数, 露出率を比較した。

表 1 がその結果である。表 1 中の重複条件はそれぞれの分野に対する最大重複数を表している。(例えば, "1122211" は最初の分野への重複条件は 1, 次の分野へも 1, 3 つ目の分野へ対する条件は 2, といった形式である。) この結果からは提案手法が以前の手法と比べ, 最大出題回数を減らし, 露出率を改善していることがわかる。

4 おわりに

本稿では露出率を最小化する複数等質テスト構成法を提案した。本手法は著者らの先行研究手法と比較し低い露出率のテスト群を構成可能である。ただし, 本手法の露出率の改善はあまり大きくない。これはおそらく, 探索アルゴリズムの効率の問題と考えられる。本手法で解く最適化問題は最大クリーク問題ではないが, 探索に使用しているアルゴリズムは最大クリーク探索アルゴリズムを強引に改変したものであり, 効率の良い探索が行われていないと考えられる。今後の課題としては, 提案手法の最適化問題にあった探索アルゴリズムを開発が挙げられる。

参考文献

- [1] Dmitry I. Belov and Ronald D. Armstrong. A constraint programming approach to extract the maximum number of non-overlapping test forms. *Computational Optimization and Applications*, Vol. 33, pp. 319-332, 2006.
- [2] Dmitry I. Belov. Uniform test assembly. *Psychometrika*, Vol. 73, No. 1, pp. 21-38, 2008.
- [3] Takatoshi Ishii, Pokpong Songmuang, and Maomi Ueno. Maximum clique algorithm for uniform test forms. *The 16th International Conference on Artificial Intelligence in Education*, 2013. to be published.
- [4] Walter D. Way. Protecting the integrity of computerized testing item pools. *Educational Measurement: Issues and Practice*, Vol. 17, pp. 17-27, 1998.
- [5] Elissavet Georgiadou, Evangelos Triantafyllou, and Anastasios A. Economides. A review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *The Journal of Technology, Learning, And Assessment*, Vol. 5, No. 8, 2007.
- [6] Hiroaki Nakanishi and Etsuji Tomita. An $o(2^{0.19171n})$ -time and polynomial-space algorithm for finding a maximum clique. *Information Processing Society of Japan SIG Technical Report*, Vol. 2008, No. 6, pp. 15-22, 2008.