

# 潜在ランク理論に基づくコンピュータアダプティブテスト

○木村 哲夫\*\*, 永岡慶三\*\*

\*新潟青陵大学看護福祉心理学部, \*\*早稲田大学大学院人間科学研究科

## 1. はじめに

潜在ランク理論 (latent rank theory, LRT) に基づくコンピュータアダプティブテスト(CAT)はまだ開発されていない。本研究において、LRTはニューラルテスト理論 (neural test theory, NTT: Shojima, 2007) のことをさす。これまでに考えられた一般的な CAT のアルゴリズムを振り返りながら、LRT ではどのようなアルゴリズムがよいか、特に項目選択ルールについて論じ、LRT に基づく CAT アルゴリズムを提案する。また、今後の研究の方向性を報告する。

## 2. 一般的な CAT アルゴリズム

コンピュータを使っていないのでCATとはいえないが、その CAT アルゴリズムの原型は、Lord (1971)の self-scoring flexilevel test にある (出題する問題の困難度があらかじめ大まかに判断されており、真ん中の困難度の問題から解答させ、解答をマークするたびに、赤か青の色が現れ、受験者がその色に従って、次に解く問題を選ぶ。受験者は最初の問題を除く設問の半分だけを解くことになる)。受験者のレベルにあった問題を解かせ、難しすぎる (あるいは易しすぎる) 問題は解かせないことで、精度を

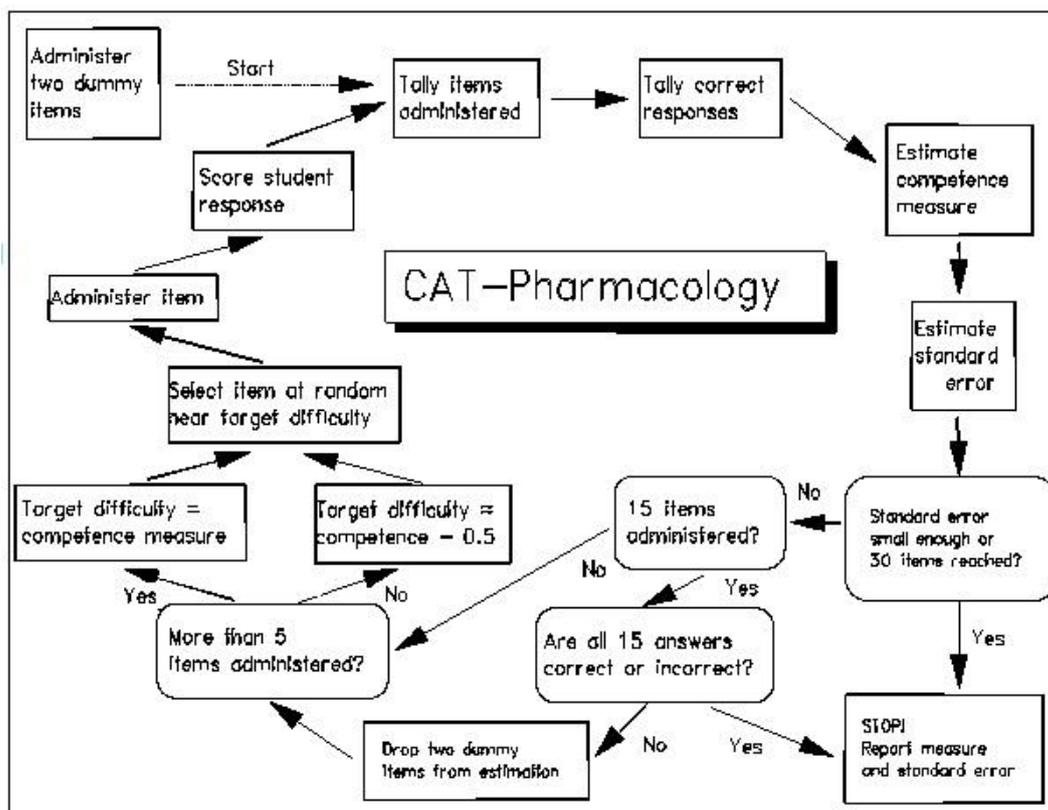


図 1 Halktis(1993)'s CAT-Pharmacology algorithm (<http://www.rasch.org/rmt/rmt64cat.htm> より引用)

落とさずに効率よく受験者の能力を推定するという、現代の CAT の目指していることが実現されているからである。

Lord(1971)の方法をそのまま CAT 化することもできる。アイテムバンクに項目数が少なく、受験させたい項目数の 2~5 倍程度しかない場合は、単純だが有効な方法かもしれない。実際、DeAyala & Koch(1986)は、Lord(1971)の方法をコンピュータで実現し、シミュレーションデータにより、flexilevel CAT が、ベイズ推定法に基づく CAT の結果と比較して遜色ないことを示し、項目応答理論 (item response theory, IRT) に基づき項目特性を求められたアイテムバンクを用意しなくても実行可能であり、教室環境で有効な方法であることを示唆している。

ただし、ここでは学部・学校単位規模の CAT を念頭に、もう少し項目数が多い場合について、Halkitis (1993) の看護学の学生の薬理学の能力を測定するテストのために開発された Rasch Model による CAT アルゴリズム(図1)を参考に、工夫すべき点をいくつか明らかにしたい。

(1) 初めにダミー項目を 2 つ置いている：項目困難度が 0 のものを 1 問正解・1 問不正解したものとして初期能力値を推定し、実際に 15 項目解答が終わるまで、毎回の能力推定値の計算にもこのダミーを含め、15 項目解答が終わったらダミーを除外する。これにより初期の緊張による失敗や偶然の正解により能力推定が大きく変わってしまうことを防いでいる。

(2) 第 5 項目までは、推定能力よりも 0.5logits 低い困難度の項目の近くからランダムに選んでいる：推定能力と同じ困難度の項目を選ぶと正答確率が 50%になるのに対して、この場合 62%になり、CAT に不慣れな学生に対して心理的負担を少なくしており、ランダムに選択することで item exposure をコントロールしている。

(3) 15 項目すべて正解あるいは不正解の場合

テストを終了し最高あるいは最低スコアを提示する。それ以外の場合は 30 項目解答するか、標準誤差が 0.4logits より小さくなった場合に終了する。

### 3. LRT に基づく CAT アルゴリズム

LRT は順序尺度上にテストを標準化する理論である点が、IRT と大きく異なる。LRT では項目の特性は項目参照プロフィール (item reference profile, IRP) で表されるが、これは他の IRT の場合(項目特性曲線を描くこと)と似ているところが多い。IRT と LRT が大きく異なるのは受験者の特性を、一義的に能力推定値をランクとしてとらえるだけでなく (LRT で受験者の潜在ランクを順序尺度上に推定することは、IRT で潜在能力を連続変数上の一つの値で推定することと同義である)、ランクメンバーシッププロフィール (rank membership profile, RMP) として多義的に表現できる点である。

このような LRT の特徴を踏まえ、また、前節でみた一般的な CAT アルゴリズムの工夫を参考に、LRT に基づく CAT がどのようなものであればよいか以下に提案する(図 2 参照)。

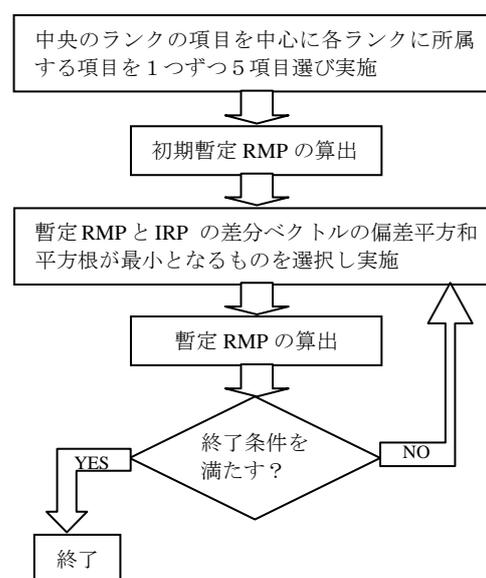


図 2 LRT に基づく CAT アルゴリズム

(1) 中央のランクの項目を中心に各ランクに所属する項目を1つずつ5項目選び実施した結果により初期暫定 RMP を算出する：項目困難度が連続尺度上でなく順序尺度上なので様々な困難度の項目を組み合わせやすい。

(2) その後は暫定 RMP と IRP の差分ベクトルの積和平均が最小となるものを選択し実施する(詳細は4. で後述)：IRT の項目選択ルールについて論じている Van der Linden (1998) の中に出てくる Maximum Expected Posterior Weighted Information という方法に相当し、おそらく、CAT の初期 (RMP ならぬ) には、全般に識別力が高い IRP を持つ項目が選択され、CAT 終期 (RMP がとがってくる) には、局所的に (受験者の暫定ランクの付近で) 急峻な IRP を持つ項目が選択されると思われる (荘島, personal communication, May 9, 2010)。

(3) 終了条件は指定項目数に達した場合か、RMP の変化が一定数以下(たとえば、0.05 以下) になった場合に終了させる。

#### 4. 暫定 RMP と IRP の差分ベクトルの積和平均による項目選択ルール

分かりにくいので、以下に例示しながら説明を加える。表1に示すような IRP が実施しようとする CAT のアイテムバンクにおいて得られているとする。

表1 IRP

ItemID	R01	R02	R03	R04	R05
Item001	0.27	0.41	0.59	0.70	0.75
Item002	0.42	0.47	0.58	0.72	0.82
Item003	0.24	0.27	0.36	0.49	0.60
Item004	0.28	0.32	0.42	0.56	0.66
Item005	0.50	0.56	0.66	0.75	0.81
Item200	0.73	0.81	0.89	0.94	0.95

\* 編みかけ部分は  $\beta$  (項目困難度指標) の位置を表す。

初期5項目を実施した後に算出された初期暫定 RMP 表2が得られた後、この RMP の形状に一番近い IRP を探す計算を以下のように行い次に出題する項目を決定する。

表2 初期暫定 RMP

R01	R02	R03	R04	R05
0.16	0.18	0.20	0.22	0.23

表3 差分 IRP

ItemID	D01	D02	D03	D04
Item001	0.14	0.18	0.11	0.04
Item002	0.05	0.11	0.14	0.10
Item003	0.04	0.09	0.13	0.11
Item004	0.04	0.10	0.13	0.10
Item005	0.06	0.10	0.09	0.06
Item200	0.08	0.08	0.05	0.01

(1) 表1からアイテムごとにランク間の差を算出し、表3のように差分 IRP を計算する。

(2) 表3の各項目の D01~D04 と表2の最初の4つの値 R01~R04 の積をそれぞれ計算し合計する (SUM1)。

(3) 表3の各項目の D01~D04 と表2の最後の4つの値 R02~R05 の積をそれぞれ計算し合計する (SUM2)。

(4) SUM1 と SUM2 の平均値を項目ごとに計算し最小の値となる項目を選択する。

アイテムバンクに蓄積されているアイテム数にもよるが、200項目以上あり、Moodle 上で複数同時受験させる状況を考えると、計算負荷がサーバーにかかり過ぎるので、なんらかの工夫が必要である (秋山, personal communication, July 20, 2011)。暫定 RMP の値が大きいランクから3番目までの  $\beta$  である項目だけについて計算を行う、暫定 RMP の最大値のランクが  $\beta$  で

ある項目だけについて計算を行う、などが考えられる。どの範囲でどの程度の規模のアイテムバンクまでなら、一般的なサーバースペックで実用的な時間内に計算できるのか、シミュレーションによる研究成果が期待される。

ある程度の幅を持たせてランダムに選択するようにしないと、特定の項目だけが何度も使用される可能性があり、item exposure の管理という観点からも、今後シミュレーションによる研究成果が期待される。

## 5. 今後の方向性について

まずは、上記の提案に基づき開発された Moodle 上で LRT に基づく CAT を実行するモジュール (秋山・木村・荘島, 2011)により、あらかじめ Exametrika(Shojima, 2008)を使い、LRT に基づき分析した英語テストのアイテムバンク (木村・永岡, 2010)の一部を使い、CAT を実施した結果を、シミュレーションデータとの比較も含め分析を行う予定である。

ひとつの文章に複数の設問が設定されている読解問題のように、大問形式の問題についても、今回の提案と同じ方向で LRT-CAT アルゴリズムを提案・開発を行いたい。

また、Linacre (1987)が提案する UCAT のように、教員の経験的判断により問題の困難度を仮に設定してアイテムバンクの中に加え、CAT をしばらく実施した後に、再計算させる方法も小規模 CAT では有益な方法である。UCAT は Rasch Model に基づく BASIC で書かれた CAT 実行プログラムであるが(プログラムを Moodle で動作する作業も進行中: Kimura & Ohnishi, 2011)、同等のことを LRT の枠組みでできるようにする仕組みについても検討を行いたい。

## 6. 終りに

とかく CAT の設計では時間の短縮と精度の

向上を目指し、項目選択において情報量を重視する傾向にある。しかしながら、特にまだ CAT 自体が教育現場に広く普及していない段階では、50%の正答確率の問題ばかりを出題する CAT は、学習者の自己効力感を削ぎ、学習に対するモチベーションを下げってしまうという負の効果を生んでしまわないだろうか。2. の (2) のように受験者の心理的側面に配慮した CAT アルゴリズムの設計を目指したい。

### 謝辞

本研究の一部は、科学研究費補助金基盤研究 (C)(課題番号: 22520590) を利用して行われました。データ分析に関して、大学入試センター研究開発部の荘島宏二郎准教授より多くの助言をいただきました。ここに心より感謝申し上げます。

### 引用文献

- 秋山實・木村哲夫・荘島宏二郎(2011) LRT モデルに基づく CAT の開発とシミュレーションによる特性解析. 日本テスト学会 第9回大会 発表論文抄録集.
- DeAyala, R. J. & Koch, W. R. (1986). A Computerized Implementation of a Flexilevel Test and Its Comparison with a Bayesian Computerized Adaptive Test. (ERIC, ED269437).
- Halkitis, P. N. (1993). A computer-adaptive testing algorithm. *Rasch Measurement Transactions* 6:4, 245-5.
- 木村哲夫・永岡慶三 (2010) Moodle による小規模 CAT 構築に向けて 1 : アイテムバンクの拡充. JSET26 講演論文集, 343-344.
- Kimura, T. & Ohnishi, A. (2011). Moodle UCAT beta version: a computer-adaptive test module based on Rasch model. *JALTCALL 2011*.
- Linacre, J. M. (1987). UCAT: a BASIC computer-adaptive testing program. MESA Psychometric Laboratory. (ERIC ED 280 895).
- Lord, F. M. (1971). The self-scoring flexilevel test. *Journal of Educational Measurement*, 8, 147-151.
- Shojima, K. (2007). Neural test theory. DNC Research Note, 07-02.
- Shojima, K. (2008-). Exametrika Ver. 4.3 [Computer software]. Retrieved July 7, 2010, from <http://www.rd.dnc.ac.jp/~shojima/exmk/index.htm>
- Van der Linden, W. J. (1998). Bayesian item selection criteria for adaptive testing. *Psychometrika*, 63, 201-216.

E-mail : kimura@n-seiryu.ac.jp  
k.nagaoka@waseda.jp