

潜在ランク理論に基づくアダプティブテストの開発と シミュレーションによる特性解析

秋山 實
東北大学大学院

1. はじめに

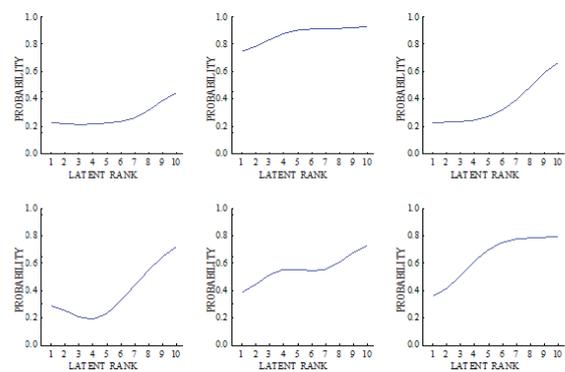
潜在ランク理論 (Latent Rank Theory, LRT) は、発表当初はニューラルテスト理論 (Neural Test Theory, NTT: Shojima, 2007a) と呼ばれていた。「潜在ランク」は離散値で表わした受験者の能力である。項目応答理論 (Item Response Theory, IRT) では受験者の能力を連続値で表わしているが、テストの解像度は連続値で表わすほど高くないため、LRT では5から20程度の潜在ランクで表わす。これによって測定が荒くなるのではなく、アイテムの特性を表わす Item Reference Profile (IRP: 図-1 参照) や受験者が各ランクに所属する確率を表わす Rank Membership Profile (RMP: 図-2 参照) といった概念の導入により、IRT よりもきまこまかくその特性を表現できるという特長を持っている。

アダプティブテスト (Computerized Adaptive Test, CAT) は、従来のテストの実施時間の半分以下の時間で実施でき、受験者の能力を考慮した出題ができることから受験者の負担を軽減できること、アイテムの露出が少なくなるため必ずしも一斉受験の必要がないことなど多くのメリットがある。CAT の動作はどの理論を用いても基本的には同じであり、受験者集団の特性、アイテムバンクの特性、初期条件、項目選択アルゴリズム、能力推定アルゴリズム、終了条件の6つの構成要素によって決定づけられる(秋山, 2010)。

2. LRT-CAT モジュール

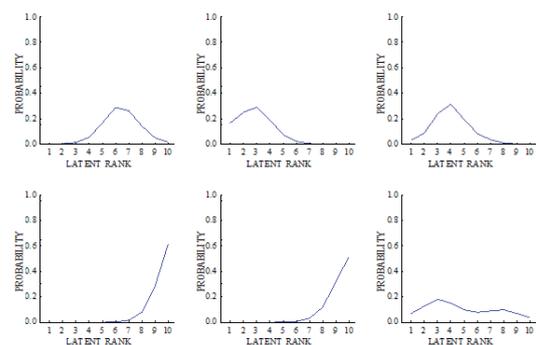
日本の大学でもっとも利用されているeラーニングプラットフォームである Moodle のプラグインとして LRT に基づく CAT 機能を LRT-CAT モジュールを開発した。LRT の特長である IRP と RMP を利用し、テストの結果もランク値だけではなく、RMP も得られ、受験者の特性をきめ

細かく把握することができる。



出典: <http://www.rd.dnc.ac.jp/~shojima/ntt/j/index.htm>

図-1 Item Reference Profile の例



出典: <http://www.rd.dnc.ac.jp/~shojima/ntt/j/index.htm>

図-2 Rank Membership Profile の例

LRT-CAT モジュールの構成要素は、以下のよう
設計した。

1) 初期設定

・初期ランク

受験者の暫定推定ランクの初期値は中心のランクとする。

・初期項目選択

最初に出題するアイテムを受験者の初期ランクとアイテムの β 値 (IRP の下のランクから正答確率を加算して

0.5を超える最初のランク:IRTの項目困難度に対応する)が等しいアイテムとする。

2) 項目選択アルゴリズム

二つのアルゴリズムからいずれかを選択できる。

・情報量最大化基準

暫定推定値としてのRMPの所属確率と出題する候補のIRPの各ランクの正答確率の積を全てのランクについて和を取り情報量とする。その最大値を持つ項目を選択する。IRT-CATではMaximum Fisher Information (, 1998)に相当するアルゴリズムである。

・初期項目選択と同じアルゴリズム

IRT-CATでは, Urry's criterion (Urry, 19)に相当するアルゴリズムである。

3) 能力推定アルゴリズム

荘島が提案したLRT版Maximum Likelihood estimation (荘島, 2007b)およびこれを情報量で重み付けをしたWeighted maximum Likelihood estimationの二つのアルゴリズムからいずれかを選択できる。

4) 終了条件判定

項目選択のアルゴリズムで使用した情報量から標準誤差を計算し、標準誤差が指定値以下(アイテムバンクの100アイテムから算出した値:最小はランダムに選択した50アイテムの情報量から算出し、最大値は情報量の多いアイテムから50アイテムで構成するテストの情報量から算出する[0.1から0.3程度])になった時に終了する。また、実用上の制約として受験項目数が50アイテムを超えた場合も終了させる。

5) アイテムバンク

Moodleの問題バンクにあるアイテムを使用し、特性パラメータは、別途、登録する。

LRT-CATモジュールの画面構成は、Moodleの小テストとはほぼ同様で、受験画面は小テストと全く同じであり、利用できる問題タイプも自動採点できる全てのタイプを網羅している。(図-3参照)LRT-CATモジュールの最大の特徴は推定された潜在ランクのほかにRMPを出力できることである。(図-4参照)

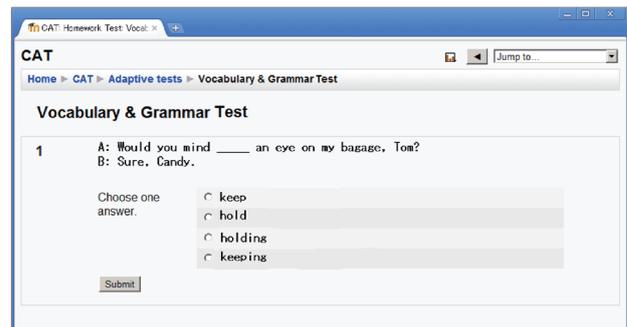


図-3 受験画面

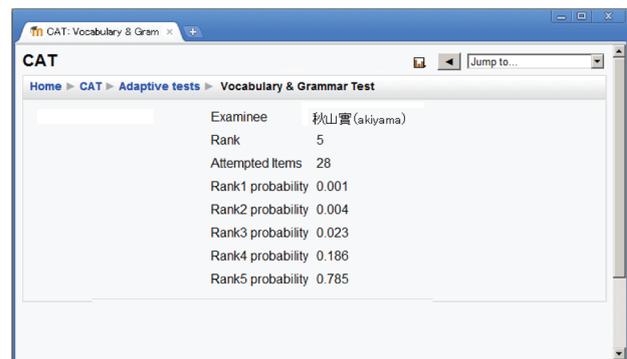


図-4 結果表示画面

3. シミュレーションの方法

CATの動作を把握する方法としては、コンピュータシミュレーションによる方法が優れている。シミュレーションでは、真値を与えてシミュレーションを行うので推定誤差を知ることができるほか、動作のトレースを取ることができる。

CATの動作をシミュレーションするには、受験者の能力(真値)と出題されるアイテムの特性値を使用し、LRTモデルを用いて正答確率を計算する。受験者の特性値は、能力ランクとRMPがあり、いずれの特性値でも正答確率を求めることができる。IRP_iを持つアイテムiに潜在ランクq_i、RMP_jの受験者jが正答する確率p(U_{ij})は次の式で表わされる。

- ・潜在ランクq_iを与える場合

$$p(U_{ij}=1) = IRP_i[q_i] \quad (1)$$

- ・RMPを与える場合

$$p(U_{ij}=1) = \sum_{q=1}^Q (IRP_i[q_i] * RMP_j[q_i]) \quad (2)$$

ここでQはランク数とする

正答確率は(1)式より(2)式の方が小さくなるが、

LRT の特長を生かし RMP を用いる方法を採用した。能力の定義から 0 から 1 までの一様乱数と比較して(2)式の正答確率が大きければ正答とする。

4. シミュレーションの結果

LRT-CAT のシミュレーション機能により生成したデータの受験者の RMP とアイテムの IRP を使ってシミュレーションを実施した。データは受験者数 $N=500$ 、アイテムは 200 アイテムである。シミュレーションの結果を表 1 に示す。(添字 e, t は推定値, 真値) 受験項目数が少なくても真値に近い推定値が得られていることが分かる。

5. おわりに

本稿では、LRT に基づく CAT を紹介した。今後の課題としては、1) 実際のテストへ適用し評価する、2) ベイズ的なアルゴリズム (EAP) を実装する、などが挙げられる。本システムは、CAT システムの構築や学習に役立つよう教育機関に限定して公開する。

引用文献

- [1] Shojima, K. Neural test theory., DNC Reasearch Note, 07-02, 2007
- [2] 秋山實, シミュレーションに基づくインハウス CAT システムの設計ツール, 日本テスト学会第 8 回大会発表論文抄録集, 148-149, 2010
- [3] Shojima K, Maximum Likelihood Estimation of Latent Rank under Neural Test Model. DNC Reasearch Note, 07-04, 2007
- [4] Urry V. W, A monte Carlo investigation of logistic test models, Unpublished doctoral disassertation, Purdue University, 1970

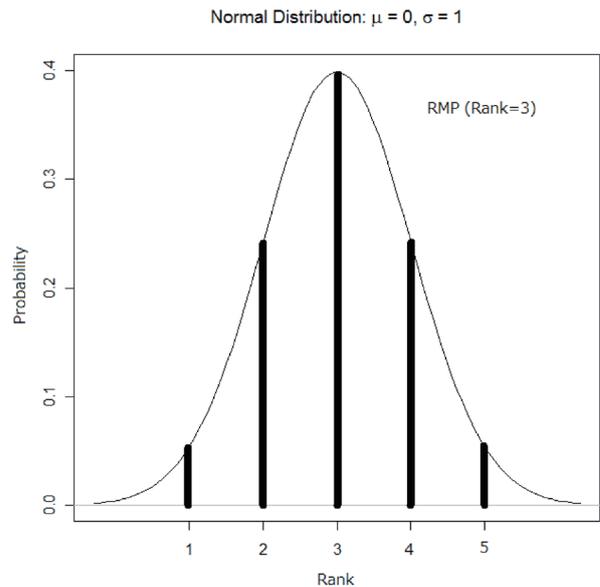


図-5 生成した RMP の例

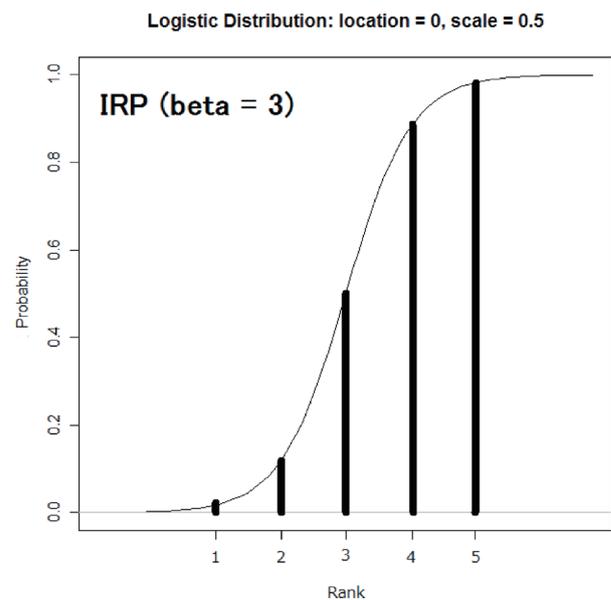


図-6 生成した IRP の例

	N		Re-Rt		RMPE-RMPt	
	mean	sd	mean	sd	mean	sd
all	11.9	5.1	0.14	0.39	0.55	0.13
rank_1	12.1	4.2	0.02	0.14	0.69	0.02
rank_2	16	5.5	0.27	0.47	0.59	0.10
rank_3	13.9	4.6	0.21	0.48	0.50	0.11
rank_4	10.6	3.4	0.22	0.50	0.50	0.13
rank_5	6.9	1.8	0.00	0.00	0.50	0.10

表-1 シミュレーションの結果の一例