

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

○秋山 實
東北大学大学院

木村 哲夫
新潟青陵大学

荘島 宏二郎
大学入試センター

1. はじめに

潜在ランク理論 (Latent Rank Theory, LRT) に基づくアダプティブテスト (Computerized Adaptive Test, CAT) を, 日本でも利用者の多いオープンソースのeラーニングシステム Moodle のプラグインとして開発した. 以後, LRT-CAT と呼ぶ.

LRT は, 以前はニューラルテスト理論 (Neural Test Theory, NTT: Shojime, 2007a) と呼ばれていた. 受験者の能力を項目応答理論 (Item Response Theory, IRT) のように連続値で表わすのではなく, ランクという離散値で表わすが, これによって測定が荒くなるというよりは, Item Reference Profile (IRP) や Rank Membership Profile (RMP) といった概念の導入により, IRT よりもきめ細かくアイテムや受験者の特性を表わす情報を表現できるという特長を持つことになった.

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ること, 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. 本稿では小規模な教育機関などで使用できる LRT-CAT プラグインとシミュレーション結果を紹介する.

2. LRT-CAT

LRT-CAT の動作は, IRT に基づく CAT の動作と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). LRT-CAT においては, LRT の特長である

IRP と RMP を利用することで IRT-CAT にはないメリットが出てくる.

LRT-CAT の構成要素は, 以下のように設計した.
「初期設定」: 最初に出題する5つのアイテムをアイテムの 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) で中心の5ランクからランダムに1アイテムずつ選択する.

「項目選択」: IRP のランク間の差分と RMP の該当するランクの所属確率の積のランクの総和 (2つの組み合わせがあるのでその平均を取る) が最も小さい項目を選択する. IRT-CAT では Maximum Expectation Posterior Waited Information (Van der Linden, 1998) に相当するアルゴリズムを採用する (木村・永岡, 2011).

「能力推定」: 荘島の LRT 版 Maximum Likelihood Estimation (荘島, 2007b) を採用する.

「終了条件判定」: 基本的には木村が提案した (木村・永岡, 2011) 「RMP の暫定推定値の変化分が規定値以下になったとき終了する」方式を採用する (本シミュレーションでは0.01とした). 後述のシミュレーション結果にもあるように受験項目数が大きくばらつく傾向があるので, 受験項目数が50アイテムを超えた場合も終了とする.

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_j*、RMP_jの受験者*j*が正答する確率*p(U_{ij}=1)*は次の式で表わされる。

- ・潜在ランク*q_i*を与える場合

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

- ・RMPを与える場合

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

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実データの受験者の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*=117、アイテムは104アイテムで、内容は英語のボキャブラリと文法のテストである。受験者1名当たり100回繰り返しシミュレーションを実施した。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)

シミュレーションの回答生成において(1)式を用いるCase1と(2)式を用いるCase2でどのような相違があるか確認するため、2つの方式でシミュレーションを行ったが、相違は認められなかった。

受験者の能力(真値)のランク*R_t*によって受験項目数Itemsに大きな差があるように見える。しかし、ランク別の受験者数に大きな差があるため、「ランクが高くなるにつれて受験項目数が少なくなる傾向にある」と指摘することとどめたい。今後、受験者の多いデータを使ってシミュレーションを行い、この点を明らかにしていく予定である。

	Items		Re-Rt		RMPE-RMPt	
	mean	sd	mean	sd	mean	sd
Case1 (N=117)	32.1	12.0	-0.07	0.61	0.31	0.20
Case2 (N=117)	32.0	12.0	-0.05	0.61	0.30	0.20
Rt=1 (N=10)	45.5	9.7	0.26	0.52	0.32	0.24
Rt=2 (N=3)	49.6	5.2	0.20	0.82	0.41	0.17
Rt=3 (N=12)	46.7	8.2	0.27	0.80	0.40	0.16
Rt=4 (N=19)	33.9	10.6	0.32	0.67	0.29	0.16
Rt=5 (N=73)	26.5	8.5	-0.25	0.45	0.28	0.20

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)同じ能力の受験者であってRMPのパターンが異なる受験者に対するLRT-CATの挙動をあきらかにする、2)IRTに基づくCATと比較する、3)実地のテストへ適用・評価する、などがある。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

- 秋山實(2010)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.
- 木村哲夫・永岡慶三(2011)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. 日本テスト学会第9回大会発表論文抄録集.
- Shojima, K. (2007a). Neural test theory. DNC Reasearch Note, 07-02.
- Shojima, K. (2007b). Maximum Likelihood Estimation of Latent Rank under Neural Test Model. DNC Reasearch Note, 07-04.
- Van der Linden, W. J. (1998). Baysian item selection criteria for adaptive testing. Psychometrika, 63, 201-216.
- E-mail : akiyama@ei.tohoku.ac.jp