

# Moodle 用プラグインの開発

## ー潜在ランク理論に基づくアダプティブテストモジュールー

秋山 實

東北大学大学院

### 1. はじめに

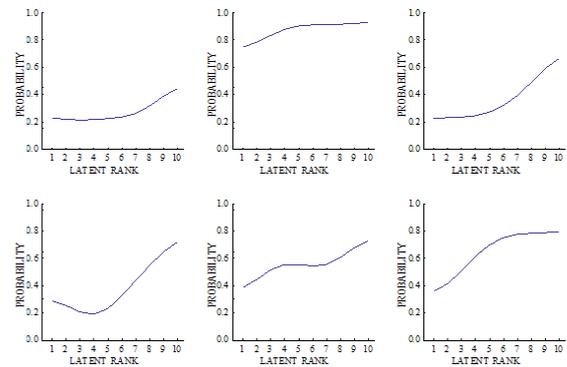
潜在ランク理論(Latent Rank Theory, LRT)は、発表当初はニューラルテスト理論(Neural Test Theory, NTT: Shojima, 2007a)と呼ばれていた。「潜在ランク」は離散値で表わした受験者の能力である。項目応答理論(Item Response Theory, IRT)では受験者の能力を連続値で表わしているが、テストの解像度は連続値で表わすほど高くないため、LRT では5 から20 程度の潜在ランクで表わす。これによって測定が荒くなるのではなく、アイテムの特性を表わす Item Reference Profile (IRP: 図-1参照)や受験者の各ランクへの所属確率を表わす Rank Membership Profile (RMP: 図-2参照)といった概念の導入により、IRT よりもそれらの特性を豊かに表現できる特長を持っている。さらに、LRT はモデルの制約が少ないため、少ないサンプル数で特性パラメータを適切に推定でき、IRTを適用できないサンプル数が少ない状況にも適用することができる(秋山, 2012)。

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ことなど多くのメリットがある。CAT の動作はどのテスト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

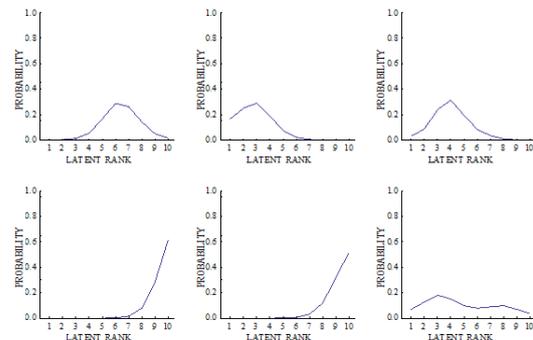
LRTに基づくCAT機能を日本の大学でもっとも利用されているeラーニングプラットフォームである Moodle のプラグインの標準インターフェース仕様に則った LRT-CAT モジュールとして開発した。LRT の特長であ

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 表示し、受験者の特性をわかりやすく提示することができる。



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

図-1 Item Reference Profile の例



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

図-2 Rank Membership Profile の例

LRT-CAT モジュールの仕様は、概ね以下のようになっている。

#### 1) 初期設定

##### ・初期ランク

受験者の RMP の初期値は各ランクとも均等に  $1/n$  (ランク数)とする。

#### 2) 項目選択アルゴリズム

二つのアルゴリズムのいずれかを選択できる。

##### ・識別力最大化基準

暫定推定ランクと隣接するランクの IRP の正答確率の

差の二乗を正答確率と誤答確率の積で除した値を識別力の指標とし、その最大値を持つ項目を選択する。IRTに基づくCATではMaximum Fisher Information (Baker, 1992)に相当するアルゴリズムである。

•LAMDA アルゴリズム

暫定推定 RMP と IRP の差分ベクトルの積和の最大値を持つ項目を選択する。Maximum Expected Posterior Weighted Information アルゴリズムに相当するアルゴリズムである(木村, 2012)。

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)である。

4) 終了条件判定

能力推定は RMP を推定することにより行うが、推定値である RMP が一定値以下の変化しなくなると同時に推定がほぼ終わったと考えられ、その時点で終了する。また、受験項目数が指定の最小受験項目数以上、最大受験項目数に達した場合に終了することもできる。

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、EXAMETRIKA が出力する EXCEL の分析結果のファイルをアップロードして与える。アイテムの対応付けは、「問題のタイトル」をキーとして行っている。(図-5参照)

LRT-CAT モジュールの画面構成は、Moodle の小テストとほぼ同様で、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)

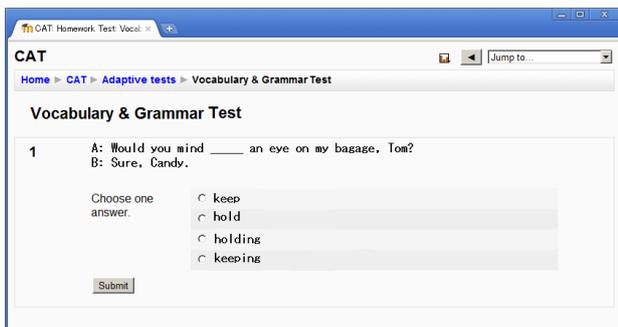


図-3 受験画面

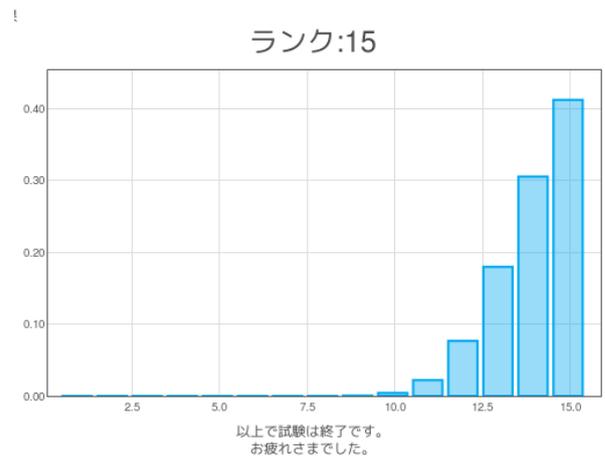


図-4 結果表示画面

3. シミュレーション機能

CAT の動作条件を決定する方法としては、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、真値が分かっているので推定誤差を知ることができる。

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<sub>i</sub>を持つアイテム i に潜在ランク q<sub>j</sub>, RMP<sub>j</sub>の受験者 j が正答する確率 p(U<sub>ij</sub>)は次の式で表わされる。

- 潜在ランク q<sub>j</sub>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、いずれの方式も選択することができる。能力の定義から 0 から 1 までの一様乱数と比較してこれからの式により得た正答確率が大い場合、正答とする。乱数を使用したシミュレーションであるので、複数回実行して誤差を算出しており、繰返し回数は 1 回から 20 回まで指定できる。推定した受験者の能力ランクが真値と一致していない推定値と全体の受験数の比率である「ランク誤差」、推定した RMP ベクトルと RMP の真値ベクトルの距離の総和を「受験数」で除した RMSD, 平均受験項目数と標準

偏差をシミュレーションの出力として表示する。(図-6 参照)

#### 4. 分散型 CAT

LRT-CAT モジュールの最大の特徴は、CAT のアルゴリズムを受験者が使用するクライアント PC 側で実行することである。このようにすることで、プレースメントテストなど一斉に大勢の受験者が受験する場合もサーバの負荷が極端に増大せず、せいぜい Moodle の標準オンラインテスト機能である「小テスト」モジュールを使用してテストを実施した場合と同等程度の負荷である。CAT のアルゴリズムは JavaScript で実装されており、FireFox3 および Internet Explorer9 に対応している。今後、スマートフォンやタブレット端末等のモバイル機器を使って受験できるよう、iOS および Android への移植作業を行っている。2013 年度には公開する予定である。



図-5 設定画面

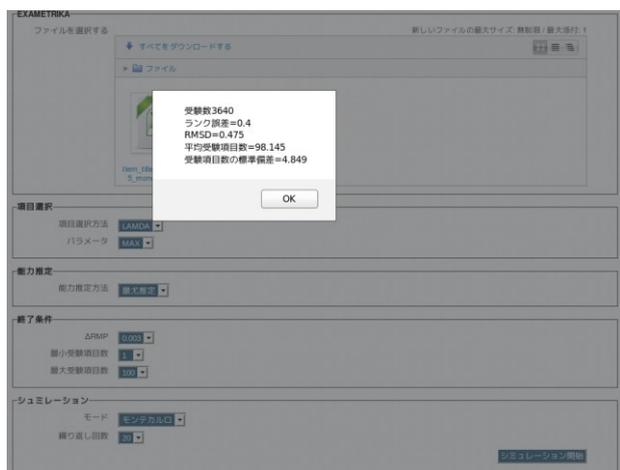


図-6 シミュレーション結果表示画面

#### 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こと、2) さらに効率的なアルゴリズムを開発・実装すること、3) タブレット端末への移植、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。

#### 参考文献

- [1] 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
- [2] 秋山實, 潜在ランク理論のパラメータ推定に必要なアイテム数と受験者数, 日本テスト学会第 10 回大会発表論文抄録集, 178-179, 2012
- [3] Baker F.B, Item response theory: parameter estimation techniques, New York, NY: Marcel Decker, 1992
- [4] 木村哲夫, 潜在ランク理論に基づくコンピュータダプティブテストアルゴリズムの提案と検証一, 日本テスト学会誌, 8, 70-84, 2012
- [5] Shojima, K, Neural test theory, DNC Reasearch Note, 07-02, 2007
- [6] Shojima K, Maximum Likelihood Estimation of Latent Rank under Neural Test Model. DNC Reasearch Note, 07-04, 2007