

# 受験者の少ないテストの時間を短縮するアダプティブテストの開発 — Moodle プラグインとしての実装と評価 —

秋山 實

東北大学大学院

## 1. はじめに

著者は、潜在ランク理論 (Latent Rank Theory, LRT) [1] に基づくコンピュータ適応型テスト (Computerized Adaptive Test based on LRT, LRT-CAT) をオープンソースソフトウェアの eラーニングプラットフォームである Moodle 2.3 のプラグインとして開発し [2], モンテカルロシミュレーションおよび実地テストで評価し、項目応答理論 (Item Response Theory, IRT) に基づく CAT が適用できない受験者の少ないテストを使って構築したアイテムバンクでも LRT-CAT が適用できることをあきらかにした。本稿では、利用者が Moodle の小テストを作成し、テストを実施した後、容易に LRT-CAT に移行できるようにするための改良を実施した結果を報告する。

## 2. LRT-CAT

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。シミュレーションによる評価の結果、適切に LRT-CAT を使うために利用者が設定すべき項目は、LRT モデルのランク数と終了条件の受験項目数であることが分かったので、その他の設定項目は固定した。

LRT-CAT の設定項目は、以下のとおりである。

### 1) 初期設定

RMP は受験者の能力確率分布である。暫定推定ランクの初期値は不明であるから各ランクとも同じ  $1/\text{ランク数}$  とする。

### 2) 項目選択アルゴリズム

暫定推定ランクにおける識別力を「隣接するランクの IRP の正答確率の差」として定義し、IRT 同様にアイテム情報量を算出し、その最大値を持つ項目を選択する。

IRT に基づく CAT ではフィッシャー情報量最大化基準 (Maximum Fisher Information criterion, MFI) 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

LRT 版最尤推定法 (Maximum Likelihood estimation, ML) [3] を著者が実装したものを採用した。

### 4) 終了条件判定

従来、推定中の RMP が変化しなくなった場合終了させていたが、受験時間にばらつきが出て運用しにくいこと、非常に少ない受験項目で終了して誤差が大きくなる場合があることなどから、受験項目数が指定の受験項目数に達した場合に終了させるよう変更した。

### 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、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。特性パラメータは、従来、利用者が手作業で受験結果をエクスポートして EXAMETRIKA で分析したのち、LRT-CAT にインポートしていたが、自己組織化マップ (Self-Organizing Map, SOM) [4] による推定機能を追加し、アイテムバンクを自動構成できるようにした。

## 3. 実行条件最適化機能

LRT-CAT の実行条件を最適化する機能を付加した。LRT モデルのランク数を 2 から 9 までの 8 通りに変化させながら、終了条件をテストを構成するアイテム数の 50% としてモンテカルロシミュレーションを実行し、各ランクの推定誤差を表示する。利用者はランク数と受験項目数を調整して、さらにシミュレーションを行い、希望するランク数かつ許容できる誤差で実行できるよう受験項目数を決定する。

## 4. 評価

モンテカルロシミュレーションで生成した回答データ

を使って評価した。理想的な特性のアイテムバンクと特性の良くないアイテムバンクを用意した。理想的な受験者特性(たとえば、LRT モデルのランク数が5の場合、能力ランク3の受験者の RMP が[0.0, 0.0, 1.0, 0.0, 0.0])、理想的な項目特性(たとえば、困難度が3の項目の IRP が[0.0, 0.1, 0.5, 0.9, 1.0])を使って回答データを生成し、これを再度、自己組織化マップ (Self-Organizing Map, SOM) で推定し、その項目パラメータでアイテムバンクを構成した。特性の良くないアイテムバンクは、たとえば、能力ランク3の受験者の RMP が[0.05, 0.2, 0.5, 0.2, 0.05]のような能力の確率分布がブロードな特性を使い、困難度3の項目の IRP が[0.3, 0.4, 0.5, 0.6, 0.7]のような識別力が低いものを使っている。このようにして生成した二つのアイテムバンクを使って LRT-CAT の性能を評価した。その結果を図-1に示す。アイテム数50の二つのアイテムバンクを使って LRT-CAT シミュレータで能力ランクを推定した。誤差は、アイテムバンクと同じ項目数50で構成したリニアテスト(全員が全ての項目を回答させる従来型のテスト)で推定される能力ランクを真値とし、受験者ごとに N=100回受験させて、真値のランクと一致しない試行  $N_e$  の比率をランク誤差(rankErr)として定義した。

$$\text{rankErr} = \frac{N_e}{N}$$

グラフの縦軸がランク誤差で、横軸は LRT モデルのランク数である。

実際のテストから構成されるアイテムバンクを使う場合のグラフは、二つのグラフの間に収まると考えられる。アイテムバンクの項目数を増やせば、誤差は比例して小さくなることも確認した。例えば、理想的な特性のアイテム150項目の場合、ランク数5でランク誤差は0.096である。

## 5. おわりに

本稿では、LRT-CAT に利用者が最適なランク数を容易に決定するため、自動的にランク数2から9でシミュレーションしてランク誤差を求める機能を付加し、その誤差範囲をシミュレーションデータで明らかにした。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予定である。

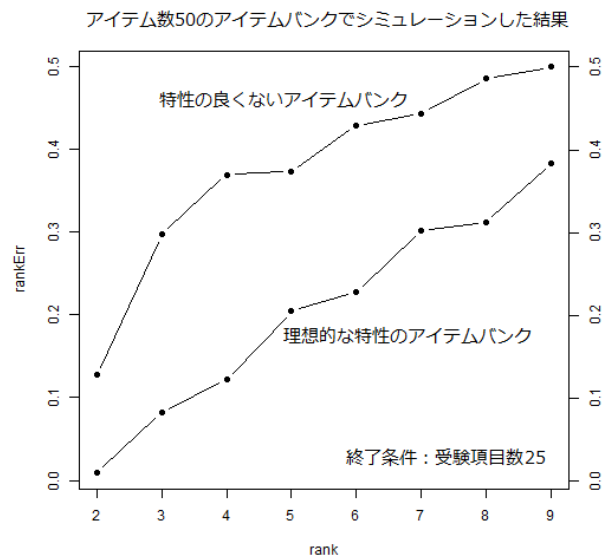


図-1 LRT-CAT の推定誤差

## 6. 引用文献

- [1]Shojima, K., Neural test theory, DNC Research Note, 07, 02, 2007, Retrieved from <http://www.rd.dnc.ac.jp/~shojima/ntt/Shojima2007R N07-02.pdf>
- [2]秋山實, 潜在ランク理論に基づく分散処理によるコンピュータ適応型テストの開発 —Moodle プラグインとしての実装—, CIEC 研究会論文誌, 4, 4-10, 2013
- [3]Shojima K., Maximum Likelihood Estimation of Latent Rank under Neural Test Model, DNC Research Note, 07, 04, 2007, Retrieved from <http://www.rd.dnc.ac.jp/~shojima/ntt/Shojima2007 RN07-04.pdf>
- [4]T. Kohonen, Self-Organizing Maps, Springer, 1997