

プレースメントテスト用アダプティブテストシステム

THE ADAPTIVE TEST SYSTEM FOR PLACEMENT

秋山 實, 東北大学大学院教育情報学教育部

Minoru Akiyama, Tohoku University Graduate School Educational Informatics

akiyama@ei.tohoku.ac.jp

概要：潜在ランク理論 (Latent Rank Theory; LRT; Shojima, 2007) は、新しいテスト理論である。この理論は、自己組織化マップというニューラルネットワークの性質を利用してテストの回答パターンを学習させ、回答のパターンが似ている受験者を同じグループに分ける。このような特性を持つ LRT はプレースメントテストに適したテスト理論である。著者は、LRT に基づくアダプティブテスト (Computerized Adaptive Test; CAT) を Moodle のプラグインとして開発し、モンテカルロシミュレーションおよびリアルデータシミュレーションで従来の項目応答理論 (Item Response Theory; IRT) に基づく CAT と同等の性能であることを明らかにした。

キーワード：潜在ランク理論, コンピュータ適応型テスト, CAT, シミュレーション, Moodle

はじめに

言語教育においては、能力別のクラス編成による教育が好ましいという考えのもとにプレースメントテストが行なわれているが、採点の手間、実施時間の制約、問題の流出リスク等を考えると、従来の全員が同じ問題を解く「紙と鉛筆」によるリニアテストより、CAT の方がメリットが大きい。しかし、CAT にも別の課題がある。現在利用されている CAT のほとんどは、IRT にその基礎を置いており、CAT に不可欠なアイテムバンクを構築することが大きな負担になっている。アイテムバンクの構築に必要なプレテストの回答データは、最も少なくても済む 1 パラメータロジスティックモデルの場合でも 20 問、200 名分のデータが必要といわれており、小規模な利用においては、取得することが難しい場合も多いと考えられる。LRT では、IRT よりモデルの制約が少ないため、IRT の 1 パラメータロジスティックモデルよりも少ないデータでアイテムバンクを構築できると思われる。

潜在ランク理論

LRT は、受験者の能力を IRT が間隔尺度の連続値で測るのに対して、5 ないし 20 程度のランクという離散値で表す。庄島は、「身長計がミリメートルの単位まで、体重計がグラムの単位までの精細な解像度を持つのに対し、100 点満点のテストが 1 点 1 点までの細かい解像度を持っているとは考えにくい。すなわち、テストは、学力をせいぜい 5~20 段階に見分けるくらいの分解能しかない。」(庄島, 2007) と述べている。しかし、一方では、受験者の特性をランクメンバーシッププロファイル (Rank Membership Profile; RMP) という受験者が各ランクに所属する確率を要素として持つベクトルで表現し、問題の特性もアイテム参照プロファイル (Item Reference Profile; IRP) という各ランクの受験者が正答する確率を要素として持つベクトルで表現するので、IRT より多くの情報が提供される。

また、IRT では、1) 受験者の能力は正規分布している、2) テストを構成する個々の問

題は独立している、という前提で成り立っているのに対し、LRTは、2)だけを前提としているので、制約が少ない分、モデルの特性を推定するために必要なデータは少ないと思われ、小規模な利用において重要な特徴となることが予想される。

シミュレーションによる最適化

LRTに基づくCAT (LRT-CAT) と項目応答理論に基づくCAT (IRT-CAT) の性能を比較して同等以上の性能を有するかどうかは、シミュレーションにより確認することができる。CATのシミュレーションには、モンテカルロシミュレーションとリアルデータシミュレーションが一般によく利用されている。モンテカルロシミュレーションは、回答をテスト理論の確率モデルにしたがって生成することにより受験者の回答行動を模擬するもので、問題や受験者の特性を自由に設定できる。リアルデータシミュレーションは、過去に実際に受験者が受験した時の回答をシミュレーションに使うもので、ある問題に対するある受験者の回答は常に一定であり、テスト理論のモデルに左右されないという特徴がある。これらのシミュレーションで異なるテスト理論に基づくCATの性能を比較することができる。

一方、小規模なCAT用の問題セット(アイテムバンク)しかない場合、CATの動作はアイテムバンクを構成する問題の特性に依存する可能性が高い。著者が開発したLRT-CATプラグインは、過去のテストの問題とその回答を利用して、アイテムバンクを構成し、さらにリアルデータシミュレーションを行う機能とモンテカルロシミュレーションを行う機能を持っており、この機能を使ってシミュレーションを行うことで小規模なアイテムバンクであっても、最適な結果を得られる設定(何問で終了させるか、どれくらいの精度を求めるか、どのようなアルゴリズムで問題を選択するか、どのようなアルゴリズムで能力を推定するかなど)を決めることができる。

シミュレーションの結果

二つのシミュレーションの結果概要は、ポスターで明かにするが、LRT-CATはIRT-CATとほぼ同等の性能を持つことが明かになった。

参考文献

- Shojima, K. (2007). Neural Test Theory. The International Meeting of the Psychometric Society 2007, Tokyo, 160.
庄島 宏二郎・植野 真臣, 学習評価の新潮流, p84, 2010