

2013年1月23日

博士学位申請論文審査報告書

大学名 早稲田大学
研究科名 人間科学研究科
申請者氏名 木村 哲夫
学位の種類 博士(人間科学)
論文題名 潜在ランク理論を用いたコンピュータ適応型テストのためのアルゴリズムの提案と実装
Proposition and Implementation of Algorithm for Computer Adaptive Test Based on Latent Rank Theory
論文審査員 主査 早稲田大学教授 永岡 慶三 工学博士(慶応義塾大学)
副査 早稲田大学教授 野嶋栄一郎 博士(人間科学)(大阪大学)
副査 早稲田大学准教授 森田 裕介 博士(学術)(東京工業大学)
副査 大学入試センター研究開発部准教授 荘島宏二郎 博士(工学)(東京工業大学)

近年、教育へのコンピュータあるいは ICT(情報通信技術)の利用は急速に進展している。特にコンピュータあるいは IT のテスト/評価への利用は、きわめて効果的な応用の一つである。コンピュータ利用をテスト/学習評価という側面に限って考えると、コンピュータによるアダプティブな出題は適応型テスト(Computer Adaptive Testing. 以下, CAT)と言われ、コンピュータ利用テスト(Computer Based Testing. 以下, CBT)の最も特徴的な機能である。それは受検者の回答履歴に応じて実時間で最適な出題を行う適応型テストを実現するものである。

本研究の目的は、理論と実践の両面からコンピュータ適応型テストについて検討を加え、新規のアルゴリズムを提案するとともに、オープンソースを利用した小規模 CAT 開発の道筋を示すことである。本研究では、1つの教育機関の1学年分の学習者を対象とする CAT 開発を念頭においており、小規模 CAT とは受験者数 200 名前後を想定している。具体的には次の3点について、理論と実践の両面から論じることである。

- (1) テスト理論の変遷と CAT の根源をふりかえり、小規模 CAT 開発に適したテスト理論と CAT アルゴリズムを検討・提案する。
- (2) CAT 開発に利用可能なオープンソースにはどのようなものがあるのか整理し、実際にそれを利用して小規模 CAT 開発を行う。
- (3) 開発したシステムを英語教育への適用することにより検証するとともに、CAT の受験結果に診断的情報を付加する方法を検討する。

適応型テストは従来、大規模な開発と実施によってのみ可能であったが、本システムにより個々の教員が簡単にパソコンを使用して実施できる使いやすさを実現した。特に CAT における出題選択アルゴリズムに新しい提案を行ったことがシステムの機能を高めた。実際、そのシステムを大学での英語教育へ適用し、潜在ランク理論(Latent Rank Theory. 以下, LRT)を用いて受検者をランク評価することで、教員側にも受検者側にも利点のある効果的なシステムであることを確認・実証した。

本研究の内容は、理論的研究と実践的研究の両面からの検討・開発からなる。それぞれにつき以

下に詳述する。

〔理論面〕 テスト理論の変遷を振り返り、CAT の根源について考えた上で、本研究で利用するテスト理論をラッシュモデル(Rasch Model. 以下、RM)および新しい段階評価に適したLRTとした。LRTに基づくCATシステム開発に関しては、次の提案を行った。

- (1) CAT を実行するためのアイテムバンクを構築する際に、望ましくない項目をLRTの枠組みで除去する指針を提案した
- (2) LRTに基づいて行うCATアルゴリズムについて、項目選択ルールと終了条件についての提案を行った

LRTに基づくCATアルゴリズムは、シミュレーション研究と実テストの両面から、さらに考察を加えることで、より機能が向上することを確認した。

適応型テストにおいて、もっとも主要な機能である識別度の高い項目を最適選択し出題するための、LRTに基づくアルゴリズムとして下に示す識別度の評価式を新たに作成した。このλ値の大きい項目を選択することで、その場その場において、識別度の高い項目を受検者に提示することが可能となる。

$$\lambda_j^{(n)} = \frac{\sum_{q=1}^{Q-1} p_{iq}^{(n)} \delta_{jq} + \sum_{q=1}^{Q-1} p_{iq+1}^{(n)} \delta_{jq}}{2}$$

ここに、Qはランク数、 P_{iq} 等は受験者iの R_q に対する暫定的な所属確率、 δ_{jq} 等は各潜在ランクの受験者の正答確率の差分ベクトルである。

〔実践面〕 大学入学の英語基礎力を測定する小規模なCATを開発するためにRMとLRTに基づいて行った一連の実践的研究から、次の5つのことを示すことができた(理論面での連番を継続する)。

- (3) 小規模CATのためのアイテムバンク構築は、RMまたはLRTに基づき行えること
- (4) 項目分析はRMについてもLRTについてもオープンソースを利用して実行できること
- (5) CATを実行するためのプラットフォームをオープンソースLMSであるMoodleを使って開発できること
- (6) シミュレーションにより用意できたアイテムバンクを使ってCATを実行した場合、どの程度の結果が得られるか、アイテムバンクのどこに弱点があるかを把握できること
- (7) 実際にCATを実施し、用意できたアイテムバンクを使ってどのレベルの項目の使用頻度が高くなるか(そのレベルの項目を今後追加すべきか)を把握できること

さらに実用面での進展は次のような事項に見いだせる。

診断的情報をCATの結果に付加する方策については、LRTにおける受検者の潜在能力のとらえ方であるランクメンバーシッププロファイル(rank membership profile. 以下、RMP)を利用することと、能力記述文(can-do statement. 以下、CDS)を使った自己評価の試みから、次の2つのことを示した。

- (8) LRTに基づき分析した場合、潜在ランクだけでなくRMPを示すことで、変化の状況をより細かく示すことができること
- (9) CDSを使って自己評価することは、自分の能力を過大評価(あるいは過小評価)していることに気づかせることができること

RMPは、受検者が結果を解釈する際や、教員(テスト実施者)が結果についてフィードバックを与える際に、大いに参考になる。一方、CDSに表されていることができるかどうか、実際にテストで測定している潜在能力とは別次元のことであり、単純に結びつけることは難しいが実データを今後広く収集していくことで、より実用性の高いシステムへ近づくことが予想される。

なお、本論文の一部が掲載された主な学術論文は以下のとおりである。

- [1] 木村哲夫：2009 ニューラルテスト理論による英語プレイスメントテストの作成と評価，関東甲信越英語教育学会研究紀要，23，23-34 頁。
- [2] Kimura, T.：2009 Construction of a Moodle-based Placement Test and Possibility of a Moodle-based Computer Adaptive Test, 全国英語教育学会紀要，20，161-169 頁。
- [3] 木村哲夫，永岡慶三：2012 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アルゴリズムの提案と検証ー。日本テスト学会誌，8，69-84 頁。
- [4] Tetsuo Kimura, Keizo Nagaoka(2011-07) Reliability of Can-Do Statements about EFL Learners, Pacific Rim Objective Measurement Symposium (PROMS) 2011, CONFERENCE PROGRAMME & ABSTRACTS, 47-48.
- [5] Tetsuo Kimura, Keizo Nagaoka(2011-10) Psychological aspects of CAT: How test-takers feel about CAT, International Association for Computerized Adaptive Testing Conference(IACAT2011), Conference (CA, USA), 47-47.
- [6] Tetsuo Kimura, Keizo Nagaoka (2012-08) Psychological aspects of CAT: seeking item selection rules which do not decrease test takers' learning self-efficacy and motivation. International Association for Computer Adaptive Testing Conference 2012 Australia Conference Program, 12-12.

以上より，博士(人間科学)の学位を授与するに十分値するものと認める。

以上