

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 10 日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2014～2015

課題番号：26750075

研究課題名(和文)簡易型項目バンクを用いたe-テストシステムの開発

研究課題名(英文)Development of e-testing system with simplified item bank

研究代表者

森 一将(MORI, Kazumasa)

東京大学・教養学部・特任講師

研究者番号：10616345

交付決定額(研究期間全体)：(直接経費) 2,300,000円

研究成果の概要(和文)：本研究では、統計的専門知識のない実務家が使いやすいe-テストシステムを開発した。このシステムは、定期的な学力テストを対象とし、項目の蓄積と特徴の類型化を行う簡易型項目バンクを中心に開発された。e-テストシステムは、理論的には多値型IRT(項目反応理論)モデルを前提としたものであり、本研究で理論的に解決された(1)多値型IRTの順序保存統計量の導出、(2)近似的なベイズ推定量の導出、(3)等化の共通被験者(項目)計画における精度高い共通被験者(項目)の選択方法、の3つの結果を基に構築されている。本研究の結果として学術論文3報、学会発表11報が得られた。

研究成果の概要(英文)：In this study, we developed e-testing system for practitioners. At first, based on polytomous item response theory model, we derived order preserving statistics and approximated Bayesian estimator. We also proposed some selection methods for equating. From these theoretical results, we developed e-testing system including simplified item bank. This system was evaluated some real datasets.

研究分野：教育工学

キーワード：項目反応理論 項目バンク 等化 テスト作問支援

1. 研究開始当初の背景

学力テストにおいて、各項目の正答・誤答反応を基に、受験者の能力や、項目ごとの反応に関する特徴を母数化するモデルをIRT（項目反応理論）モデルという。特に、1つの問題が多段階の得点を持つ形式のものを多値型IRTモデルという。本研究では多値型IRTモデルの母数推定に関して次の3つの点に着目する。

(1) IRTモデルの母数推定の困難さと母数順序性

多値型IRTモデルによる母数推定は、ソフトウェアによる複雑な計算が必要になり、統計や数値計算の専門知識がないと行うことができない。申請者らは、この母数推定計算の困難さを解決するために「**母数順序性**」という性質を利用することを提案している。(森,大森 2012、森,2013) 母数順序性とは、テスト受験者ごとの正答項目数や項目ごとの正答者数といったデータに簡単な計算を施し、IRT母数推定値の大小関係と同じ順序を持つ統計量を作り出す問題である。これについては以下の課題が指摘できる。

この統計量の順序性の一致は漸近的に成り立つ。つまり、受験者が多くないと順序性は一致しない。したがって、受験者数とIRT母数推定値 統計量間の順序の一致度の関係性を明らかにする必要がある。

(2) 項目バンクと類型化、等化

すでに実施されたテストの項目母数を推定し、それを基に項目を類型化したものを項目バンクという。項目バンクは新規のテストを作成するとき、項目を再利用する目的で作られるが、以下の2つの問題が生じている。

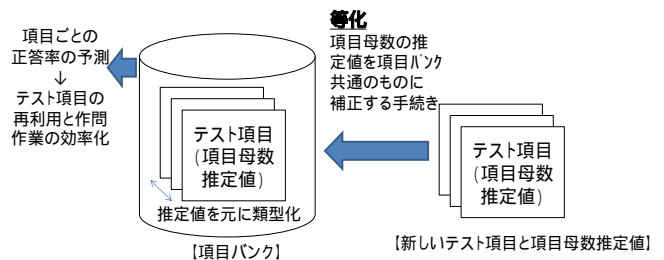
項目バンクの等化の問題。等化とは、項目バンクに対し、新たなテスト項目を追加するときに行う手続きで、項目母数の推定値を項目バンク共通のものに補正する役割を持つ。この手続きは専門知識が必要となり、実務家に対して利用の壁となっている。

項目母数の類型化の方法。例えば、前述のRSMの項目母数は困難度と呼ばれ、この推定値が大きくなると正答率が下がる。ところが困難度推定値の読み取りと、類型化の実施には、知識と共に分析の経験が必要になり、実務家が項目バンクを利用する妨げとなっている。

(3) 母数順序性と正答率

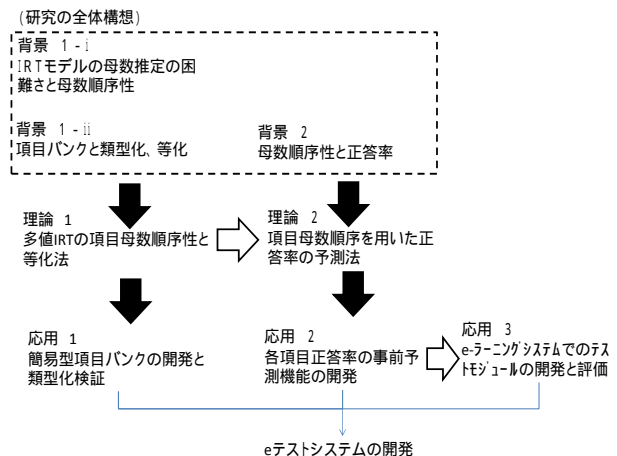
母数順序性を基にした正答率の予測。テストの作問に項目バンクを用いる利点は、IRTモデルにより作成したテ

ストの正答率を事前予測できることである。ところがこの作業には、専門的知識を要する母数推定が必要となる。この問題の解決策である母数順序性問題においては正答率の予測は考えられておらず、項目バンクへの母数順序性問題の適用は未だに十分に行われているとは言えない。



2. 研究の目的

1節で挙げたような背景を基に本研究は以下のような結果を明らかにすることを目的に行われた。



(1) (理論1) 多値型IRTの項目母数順序性成立条件の導出と等化法

申請者の先行研究を基礎として、実用的な母数順序性を保証する受験者数を検討する。

母数順序性を保証できる受験者数の基準値。シミュレーション研究などにより項目母数推定値の順序性が真値と一致する受験者数を明らかにし、基準値とする。

等化法の改善。項目バンク-新テスト間で一部を共通項目にする方法を前提とし、母数順序性に適用させる方法の提案を行う。実データなどを基に妥当性を検証する。

類型化カテゴリ数と項目数の関係の分析。適切な類型化のカテゴリ数は項目母数推定値のばらつき度合いや項目数に依存する。したがって、項目数とカ

テゴリ数の関係をシミュレーション研究を用いて明らかにし、適切なカテゴリ数についての基準を作る。

(2)(理論2) 項目母数順序を用いた正答率の予測法

予測法の提案と予測精度の検証。項目母数順序を用い、受験者集団の正答率を予測する方法を提案する。実データとシミュレーション研究で予測精度の検証を行う。

(3)(応用1) 簡易型項目バンクの開発と類型化検証

～ の理論的結果を基に簡易型項目バンクを開発する。項目バンクは多値型IRTモデルの1つであるRSMを前提とする。簡易型項目バンクでは、理論研究で得られた基準値等を用い、類型化や等化の手続きを自動的に行うよう機能を実装する。加えて実データを用い、カテゴリ内の項目母数推定値のばらつきなど分析し、従来の項目バンクとの差異を明らかにする。

(4)(応用2) 各項目正答率の事前予測機能の開発

で提案した予測法を用い、正答率の事前予測機能を開発する。この際、実務家に意見を聞き、分かり易い予測正答率の表示方法を合わせてデザイン化する。

(5)(応用3) e-ラーニングシステムでのテストモジュールの開発と評価

(応用1)の簡易型項目バンク、(応用2)の正答率予測機能を既存のe-ラーニングシステムに実装する。対象システムは広く流通し、汎用性の高い機能を持つ「moodle」を前提とする。この結果をe-テストシステムと呼ぶ。このシステムを外国語教育における小テストなどに適用し、操作性、作業の効率化、教育効果の向上の有無などの総合評価を行う。

3. 研究の方法

研究の方法は、主に2節で挙げた理論研究を2014年に実施し、その後応用研究を2015年に実施した。特に2015年度は、英語文法テストや「ビジネス数学検定」といった既存の資格試験のデータを用いて項目バンク作成の可能性を検証した。

4. 研究成果

主な研究成果は2節の(理論1)と(応用1)そして(応用3)の一部について得られた。正答率の予測に関する研究である

(理論2)と(応用2)については、当初想定していた以上に順序保存性の問題や等化法の問題の解決に時間を費やしたため、対応ができなかった。詳細は以下の通り。

(1)(理論1) 多値型IRTの項目母数順序性成立条件の導出と等化法

まず、前段階として、非常によく用いられる2値型IRTであるラッシュモデルの周辺化最尤推定量について、項目母数順序を保存する統計量(順序保存統計量)を導出した。(雑誌論文)

この結果を基に多値型IRTの能力母数と項目母数の結合最尤推定量に対して順序保存統計量を導出した。多値型IRTは、尤度関数の関数形が複雑であるため、順序保存性の問題は議論できない。そこで、本研究では、この尤度関数に1種の近似を施し、順序保存に関する計算を容易にするような関数(近似関数)を提案した。この近似関数はSpech(1960)の結果から、通常の尤度関数との差異比率の上限が明らかになっており、数学的な扱いが容易である。この近似関数を用い、RSMおよびLinear Rating Scale Model(LRSM), Partial Credit Model(PCM), Linear Partial Credit Model(LPCM)について順序保存統計量を導出した。加えて、小規模なテストの実施範囲(項目数30以下、カテゴリ数11以下、受験者数100人以下)で提案した順序保存統計量が真値と相関高く推定され、かつ近似尤度関数も最も近似が悪くなる場合と比較して差異比率が非常に低く、精度のよい近似を行うことをシミュレーション研究によって検証した。(雑誌論文)

また、本研究で提案した近似尤度を用いて、実際にベイズ推定量を導出し、これを基にした等化項目・受験者の選択法や等化法も提案した。

具体的にはRSMと近似尤度を用い、損失関数が2乗損失の場合におけるベイズ推定量を導出し、これを基にした推定量を提案した。このベイズ推定量は項目反応行列から計算した統計量の線形結合になっている。この推定量と真値との精度はクロンバックの係数として評価できるが、前述の通りベイズ推定量は項目反応行列の線形結合であるため、この評価は項目反応行列を基にした特定の関数の挙動を調べればよいこととなる。具体的には、共通受験者計画の場合には、それぞれの受験者に対し、項目の得点の分散が大きい受験者を選択し、共通項目計画の場合は、それぞれの項目に対し、受験者の得点の分散が大きい項目を選択すればよい。このことにより、共通受験者(項目)計画において、ある先行テストが実施されたときに共通項目(受験者)を選択する方法が開発された。

この選択法の利点は2つある。1つは統計パッケージによる項目・能力母数の推定を必

要とせず非常に容易であることである。もう1つは、共通受験者(項目)の選択を、従来のように問題の内容の定性的分析や受験者の過去の成績などの外部情報に依存することなく、クロンバックの係数という理論的に根拠がある指標に基づき行得ることができることである。(発表、)この結果については、現在関連する結果を取りまとめ、現在投稿準備中である。

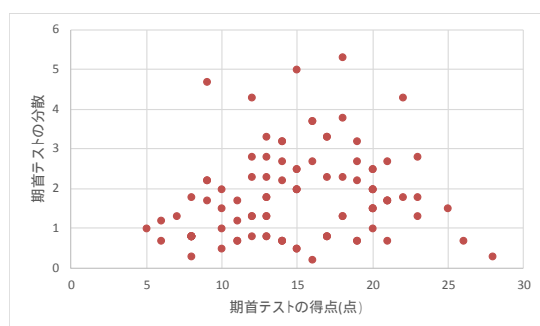
(2)(応用1)簡易型項目バンクの開発と類型化検証

(理論1)の成果のうち、ベイズ統計量と等化法を用い、簡易型項目バンクの提案を行った。この簡易型項目バンクについては、当初はシステムでの実装を予定していたが、(理論1)で得られた結果が当初の想定以上に簡略化され扱いやすいものであり、MS-EXCELなどの市販の表計算ソフトウェアで簡単に計算できるものとなったため、計算法を提示するのみにとどめた。また、等化精度をシミュレーション研究によって検証した。その結果、小~中規模(項目数50以下、カテゴリ数7以下、受験者数100人以下)の実施範囲において適切な等化係数の推定が行われることがシミュレーション研究によって確かめられた。(発表)

(3)(応用3)e-ラーニングシステムでのテストモジュールの開発と評価

(応用1)で述べたMS-EXCELを用いた簡易型項目バンクを実際のデータを使って構築し、その評価を行った。評価の対象になったデータは、大学生の英語文法に関するテストデータ(論文)と、公的な資格試験である「ビジネス数学検定」のテストデータである。このうち、本報告ではビジネス数学検定について共通受験者・項目の選択に関する分析結果を示す。(発表)このテストは、作問者(日本数学検定協会)が設定した5つのカテゴリ(把握力、分析力、選択力、予測力、表現力)に分類される。

まず、共通被験者計画を想定した時の簡易型項目バンクをMS-EXCEL上に構築し、共通受験者に適する受験者の挙動を分析した。下図がその結果である。



共通受験者計画の場合は、受験者ごとの項目の得点の分散が大きい受験者を共通受験者として選択すればよい。この結果からは共

通受験者として適するものは(期首)テストの得点分布の中央近くに多く存在することが分かる。つまり、この結果は、共通受験者は全体の受験者の中で「中程度」にできる受験者から選択すれば、良い等化が行われることを示唆している。

次に共通項目計画を想定し分析を行った結果を示す。下表がその結果である。

問題カテゴリ	受験者得点の分散	分散の順位
把握力	65.14	5
分析力	190.81	1
選択力	101.48	4
予測力	114.81	2
表現力	108.48	3

共通項目計画の場合は、項目(問題カテゴリ)ごとの受験者の得点の分散が大きい項目を共通項目として選択すればよい。このとき、望ましい共通項目として選択されたものは「分析力」や「予測力」といった基本的な数学能力を測る項目であった。本事例においては項目数が少ないため強い結論を導くことはできないが、この結果からは、共通項目の選択方法として、より基本的な内容を扱った項目を選択すればよいことが示唆された。

<引用文献>

Specht, W. (1960). Zur theorie der elementaren Mittel. *Math. Z.*, 74,91-98.

5. 主な発表論文等

[雑誌論文](計3件)

Shintani, M., Mori, K. and Ohmori, T., Image Schema-Based Instruction in English Grammar. To appear in JALT conference proceedings(JALT2015), 2016年, 査読あり.

Mori, K., The order-preserving properties of estimates in polytomous item response theory models with approximated likelihood functions. *Scientiae Mathematicae Japonicae*, 査読あり, 2015年, 2015-46.

Mori, K., The order-preserving properties of the Rasch model in marginal maximum likelihood estimation. *Scientiae Mathematicae Japonicae*, 査読あり, 2015年, e-2015, 2015-5.

[学会発表](計 11 件)

Shintani, M., Mori, K. and Ohmori, T. Image Schema-Based Instructions on English Grammar. 全国語学教育学会 41 回大会、グランメッセ静岡(静岡県静岡市) 2015 年 11 月 20 日.

森一将、非対称損失を用いた多値型 IRT 予測量の近似、統計関連学会連合大会 (2014)、東京大学(東京都文京区) 2014 年 9 月 13 日.

大森拓哉・森一将・新谷真由、LLTM による英文法問題項目分析の実践的研究、日本テスト学会第13回大会、関西大学(大阪府吹田市) 2015年9月10日.

森一将・大森拓哉・新谷真由、多値型 IRTの近似ベイズ推定量を用いた項目等化について、関西大学(大阪府吹田市) 日本テスト学会第13回大会、2015年9月10日.

森一将、多値 IRT の条件付き最尤推定値の順序保存について、統計関連学会連合大会 (2015)、岡山大学(岡山県岡山市) 2015年9月6日.

森一将、Rating Scale Model の条件付き最尤推定値の順序保存について、日本行動計量学会第43回大会、首都大学東京(東京都八王子市) 2015年9月1日.

森一将、多値型 IRT の予測量の近似について、第 42 回日本行動計量学会回大会、東北大学(宮城県仙台市) 2014年9月2日.

大森拓哉・森一将、極端な値を用いた項目バンクの等価係数の影響について、第 12 回日本テスト学会大会、帝京大学(東京都八王子市) 2014 年 8 月 30 日.

森一将・大森拓哉、多値型 IRT の近似予測量を用いた項目バンクの検討、第 12 回日本テスト学会大会、(東京都八王子市) 2014 年 8 月 30 日.

Ohmori, T. and Mori, K., A new representation method of mind image data and its applications to career selection. International Congress of Applied Psychology 2014. Paris(France), 2014 年 7 月 8 日.

Mori, K. Simple prediction methods of item parameters in polytomous IRT model. International Congress of Applied Psychology 2014. Paris(France), 2014 年 7 月 8 日.

6. 研究組織

(1) 研究代表者

森 一将 (MORI, Kazumasa)
東京大学 教養学部 特任講師
研究者番号: 10616345