

## 科学研究費助成事業 研究成果報告書

平成 30 年 6 月 8 日現在

機関番号：32641

研究種目：若手研究(B)

研究期間：2015～2017

課題番号：15K21379

研究課題名(和文)最適テスト法と双方向 e-learning システムの実装

研究課題名(英文)Implementation of optimal test method and bi-directional e-learning system

研究代表者

作村 建紀(SAKUMURA, Takenori)

中央大学・理工学部・助教

研究者番号：50735389

交付決定額(研究期間全体)：(直接経費) 2,700,000円

研究成果の概要(和文)：本研究では、テスト理論に基づく e-learning システムを構成する要素の初期条件の違いによるユーザ能力の推定精度への影響を明らかにした。これにより、推定量としてベイズ推定量の期待事後平均が安定して良い結果を示すこと、推定量の簡易なバイアス補正法の提案などの結果を得た。また、より多くの情報を活用できる新たな評価モデルの提案を信頼性工学分野の知見によって行った。本研究の成果として、学術論文5報、学会発表12報が得られた。

研究成果の概要(英文)：In this research, we clarified the influence on the estimation accuracy of user ability by the initial condition difference of the elements constituting the e-learning test system based on test theory. As a result, the posterior mean of the ability was stable and good results, and the simple bias correction method of the estimator was proposed for each ability estimate methods. In addition, we proposed a new evaluation model that can utilize more information according to knowledge in the field of reliability engineering. As a result of this research, 5 academic papers and 12 presentations of academic societies were obtained.

研究分野：計算機統計学

キーワード：適応型試験 項目反応理論 e-learningシステム 信頼性モデル

## 1. 研究開始当初の背景

少子高齢化の加速と高等教育進学率の増加を背景に学生の習熟度は多様化している。そのため、試験の結果から問題項目に潜む特性を決定づけ受験者の能力をより公正に評価する試験評価技術の確立が強く望まれている。実用的な能力評価システムの構築には、IRT(項目反応理論)(参考文献 )が有用である。IRTを用いたe-learningシステムでは、受験者の能力に合わせた適切な問題項目を出題し続けることで効率的な能力評価を可能にする。IRTを用いたe-learningシステムでの出題項目は事前の予備テストから項目特性を推定する。これにより、システムの評価性能を高めることができる(参考文献 )。個々に合わせた教育システムとして、e-learningによる学習支援は注目される一方で、適切な問題項目の評価に困難を要する。項目特性を正しく評価出来ていない場合、それに基づき推定される能力の評価もまた、精度が悪くなる。

IRTを用いたe-learningシステムの実用に成功している例は、TOEICや語学力検定などの短時間回答できる試験が多い。これは多肢選択式による回答の容易さと、試験問題数の多さに起因する。逆に数学試験のように記述式回答が主の場合、1項目あたりの回答時間が長くなり、1回の試験問題数は少なくなるため、システムの能力評価性能は落ちる。そこで、1項目の回答に対して暗に正解不正解とするのではなく、部分得点や回答するまでの反応時間などを構造に反映したIRTの応用モデルに関する研究が国内外で盛んに行われるようになり、これまでに、段階反応モデル(参考文献 )に代表されるような数多くの応用モデルが考案されてきたが、それらのモデルは複雑で高次元なものとなり、その実用例などはあまり提案されていないのが現状である。

システム受験者の回答結果は回答に要した時間や回答順序などの多くの情報を含む一方で、回答項目は受験者間で異なり、その試験結果からは項目特性の推定が困難であるため、項目特性は予備テストの結果に依存しやすい。また、予備テスト用の被験者集団の確保にも多大な労力を要するという問題もある。特に、数学試験のような1問あたりの問題に多くの時間を費やすケースでは、この問題はより顕著になる。つまり、予備テスト受験者への依存度を出来る限り軽減し、かつシステム受験者の回答結果を反映させた評価法の確立が必要となる。

申請者はこれまでにIRTを拡張し、学習支援システムでの能力評価と同時に問題項目も再評価するテスト法を提案してきた。これはシステム-ユーザー間で双方向の情報伝達を行う成長型システムであるが、実用化には初期条件の設定や適切な評価モデルを構築する問題がある。つまり、本システムにおける最適な問題項目の評価法の提

案と新たな能力推定モデルの構築を検討する必要がある。

## 2. 研究の目的

1節で挙げた背景と、e-learningシステムにおける評価精度の向上および運用コスト削減を目的の主軸とし、より妥当な評価モデルの提案、問題項目出題方法の検討による評価精度向上、不完全データの欠損値予測による項目特性の研鑽というアプローチを並行して研究してきたこれまでの結果を踏まえ、本研究では、予備テストそのものをシステム内部に組み込み、システム受験者自身によって項目特性を再評価することで、予備テストの依存性および労力を解決すると同時に逐次的に成長していく双方向e-learningシステムの構築を主目的とした上で、以下を目的に行われた。

### (目的1)e-learningシステムの初期条件の違いによる、各能力評価モデルの評価精度の検証

欠損値予測法によって可能となる事前の予備テストを行わない場合のe-learningシステムにおける能力評価精度について詳細に検討する。IRTによるe-learningシステムでの能力評価は、以下の要素に影響を受ける。

#### (1) 項目のデータベース

極端な場合として、すべての項目の項目特性が明らかになっている理想的な状況では適切な評価が可能だが、困難度レベルのような離散的な項目特性しか分からない不十分な状況では精度が落ちると考えられる。

#### (2) 出題項目を選択する出題方式の基準

出題方式として、情報量基準による方法や最小二乗誤差の方法など、推定された能力と項目特性の間の距離によって出題項目を選択するか、全くのランダムで選択するかで基準で異なる。

#### (3) 項目特性と能力の推定量

項目特性と能力の推定において、IRTでは2母数ロジスティックモデルを用いることが多い。このとき、モデリングのアプローチとして、頻度主義者の立場で見ると、周辺最尤推定量が最も有名である。または、バイアスを考慮した重み付き最尤推定量も提案されている。これらの推定方法としては、EMアルゴリズムによる方法が提案されている。一方で、ベイジアン立場で見ると、期待事後平均や最大事後モードを用いることができる。推定方法としては数値積分を近似的に行う区分求積法による方法とMCMCによる方法が提案されている。

これらの違いを含めて、能力評価精度への影響をシミュレーションベースで検証する。

## **(目的2) 学習評価モデルへの援用を期待した信頼性工学分野における信頼性モデルの提案**

より多くの情報を含める新たな評価モデルを、信頼性工学分野の知見を援用し提案する。特に、試験問題に取り組む際の受験者の回答時間に注目する。問題を見てから回答に至るまでの経過時間は、能力を評価する上で重要な情報を含んでいると考える。試験の結果として観測される正答・誤答の2値は、思考から回答までの到達後の結果と考えれば、時間とともに回答に至るプロセスは、確率過程として捉えられる。このとき、回答に伴うストレスが疲労という形で蓄積していく状態を時間に関してモデリングすることを考える。

## **(目的3) e-learning システムの構築**

実際に e-learning システムを構築し、現実の場面における問題点を把握する。

## **(目的4) IRT モデル群の再評価**

(目的1) ~ (目的3) をもとに IRT モデル群の再評価を、機械学習分野の知見を取り入れて行う。

## **(目的5) IRT モデルの他分野への応用**

教育分野と信頼性工学および機械学習分野の融合により、これまで心理・試験データに対して用いられてきたあらゆる IRT モデルは、他分野への応用が現実のものとなりうるため、それにも取り組む。

### **3. 研究の方法**

2 節で述べた研究目的に基づき、以下のように行った。なお、従来のシステムの問題点を明らかにする研究である e-learning システムの構築(目的3)と、IRT モデル群の再評価(目的4)は、e-learning システムの評価精度の検証(目的1)、信頼性工学分野の知見による問題解決(目的2) および他分野への応用(目的5)に当初予定していた以上の時間を費やしたため、対応できなかった。

### **(1) 目的1の研究手法**

事前に予備テストを行わない場合における e-learning システムの能力評価精度を検討する。つまり、出題される項目特性は不十分である状況を踏まえ、システム受験者にとって適切な項目が出題されていない状況で試験を行った場合、受験者の能力評価をどの程度正しく評価できるのかを明らかにする。具体的には、以下の通り行った。

- (1) 項目のデータベースを理想的な状態と離散的な状態の2つを用意する。理想的な状態では、項目特性は真値が明らかであるとし、適切な項目が実数空間上に無数に存在すると仮定する。離散的な状態では5段階の困難度に離散的に属するとする。また、

項目数は無限に存在するとする。

- (2) 出題方式として、情報量基準による選択方式と、ランダムに選択する方式で比較する。
- (3) 能力推定方法として、周辺最尤推定量、バイアス補正した重み付き最尤推定量、ベイズ期待事後平均、ベイズ事後モードを採用する。なお、ベイズ推定においては、MCMC による方法も別途行う。

これらの条件の組み合わせのもとで、様々な能力を想定した受験者による試験をコンピュータ・シミュレーションで検証する。

### **(2) 目的2の方法**

通常の IRT では、受験者の能力は試験を通して常に一定であると仮定されている。紙面上の試験ではその試験結果は静的なデータとして得られるが、オンライン学習のようにリアルタイムに回答結果を把握できる場合、その回答結果は前項目からの動的なデータとして得られる。信頼性工学分野においては、電気材料のストレス耐性は時間とともに変化すると考えることもできる。オンライン学習においても、回答というストレスによる疲労の一部が回答毎に蓄積し、受験者の能力を減衰させると捉え、より柔軟なモデルを構築する。

### **(3) 目的5の方法**

他分野への応用として、野球データへの IRT の適用を考える。近年、打者と投手の1球ごとの観測データを取得することが可能になっている。ここで、打席ごとの対戦データに注目すると、これはヒット・アウトの2値データとして見れば、IRT モデルの適用可能なデータとなる。このとき、打者は受験者能力、投手は項目難易度と対応できるが、その逆の見方も可能である。このとき、モデルを1母数ロジスティックモデル(ラッシュモデル)とすれば、解釈を容易にするだけでなく、オッズ比を用いた能力比較も容易になる。

### **4. 研究成果**

主な研究成果は2節の目的1、目的2、目的5の一部について得られた。

### **(1) 目的1の研究成果**

まず、e-learning システムによる試験を行った場合の学習者の能力の推定精度について、1) 理想的な項目データベースの環境下での推定方法の違いによる能力推定値と真の能力値とのバイアスを検討し、また2) 有限な出題数における学習者の能力ランクの判別に必要な問題数について明確にした。IRT を利用した e-learning システムによる試験は、学習者の能力に常に適した問題項目を出題することをコンセプトとして持つが、学習者の真の能力は未知であるため、推定された

能力に合わせた問題項目を出題するように構築される。よって、出題数の少ない序盤では能力推定値の真の能力値との誤差が生じるため、この研究成果により、序盤での評価の信頼性を明確にできた。(学会発表)

次に、能力評価モデルの推定精度の比較およびそのバイアス補正法の提案について行った。この研究では、オンライン学習における能力評価モデルに対し、同条件の試験条件下で、ある一定の能力評価精度を保つまでに必要な問題出題数を、モンテカルロ・シミュレーションによって明らかにした。シミュレーションでは、擬似的な試験を行い、各能力評価モデルで出題ごとに能力推定を行う方法に依った。この検証により、それぞれの能力評価モデルの特徴が明らかとなり、期待されていた重み付き最尤推定モデルよりもベイズ推定モデル(特に事後平均によるモデル)の方が安定した結果を示すことが分かった。MCMC モデルも同程度の水準で良い結果であったが、十分な精度を保つための計算時間が大きく、オンライン学習時の実使用場面を考慮すると、適切ではない結果となった。また、これらの結果から、能力推定精度と出題問題数との関係が明らかになった。さらに、それぞれの評価モデルにもとづいてそのバイアス補正を行う方法を考案した。その補正法を用いることで、オンライン学習の初期段階に大きく現れる能力の過大評価や過小評価を改善し、より適切に問題項目を出題するシステムとなる。(雑誌論文、学会発表)

## (2) 目的2の研究成果

信頼性工学分野のモデルとして、時間とともに過去の回答によるストレスの一部が蓄積する信頼性モデルを考案し、そのパラメータ推定には MCMC を用いる手法について提案した。このモデルは疲労が蓄積することを表現したモデルであり、一時的な疲労の回復を考慮したものになっている。これを学習時の疲労と関連付けると、e-learning 試験を受験している際に、回答ごとに回答者への負荷が逐次的に蓄積していくとともに、問題と問題の間で少しの休息がある状態と同じである。回答者は常に一定の能力を発揮すると考えるよりも、疲労とともに本来の能力が減衰していくと考えるほうが自然であるとの観点に立つと、この考案したモデルはより妥当であると考えられる。これは学習者の能力推定値に対する学習時の疲労などの影響を考慮したモデルへの足がかりとなる。(学会発表、学術雑誌)

## (3) 目的5の研究成果

能力評価を学力試験のような対象から派生させ、スポーツへの拡張を行った。具体的には、野球の打者および投手の対戦データをもとに、打者と投手の能力を同時に評価することで、従来の打者の打率や投手の防御率などの指標よりもより詳細な能力の解釈を

可能とした。本研究では、従来の項目反応理論での対象としての項目を投手に対応させており、受験者を打者に対応させている。投手(項目特性)の評価による打者(能力)の評価への影響も定量的に解釈される。この一連の研究により、被験者と項目という2者の関係は、相互に入れ替え可能であることの一例であり、双方向 e-learning システムの前提条件を支持する結果を得た。(学術雑誌)

その他、能力バイアスの発生を考慮した、新たな IRT モデルの考案の足がかりとして、ベイズ統計の知見を取り入れた研究を行い、リスク評価の観点から推定精度を調査した。(学会発表)

また、バイアス補正技術の派生研究として、室内の複数センサの情報から室内位置を、バイアスを考慮して特定する技術の開発を行った。(学術雑誌)

さらに、ガンマ分布のパラメータ推定におけるバイアス補正の研究を行った。(学会発表)

## [参考文献]

- Lord, F. M., A theory of test scores, *Psychometric Monograph*, No.7, 1952.  
Samejima, F., Estimation of latent ability using a response pattern of graded scores, *Psychometrika Monograph Supplement*, 34(4), 100-114, 1969.  
Wainer, H., Dorans, N. J., Green, B. F., Steinberg, L., Flaugher, R., Mislevy, R. J. and Thissen, D.: Computerized adaptive testing: A primer, *Lawrence Erlbaum Associates, Inc*, 1990.

## 5. 主な発表論文等

### [雑誌論文](計 5 件)

Yoshida, A., Sakumura, T. and Kamakura, T. (2017) "Indoor Location Estimation Based on TOA Data and Bias Estimation Using Gamma Regression," *Journal of the Japanese Society of Computational Statistics (JJSCS)*, Volume 30, Issue 1, 査読あり

Abe, K., Sakumura, T. and Kamakura, T. (2017) "Measurements of Baseball Players' Batting Abilities," *Proceedings of the Institute of Statistical Mathematics*, Volume 65, Number 2, Pages 235-249, 査読あり

Sakumura, T. and Hirose, H. (2017) "A Bias Reduction Method for Ability Estimates in Adaptive Online IRT Testing Systems," *International Journal of Smart Computing and Artificial Intelligence, International Institute of Applied Informatics*, Volume 1, Issue 1,

Pages 59-72, 査読あり

Sakumura, T. and Kamakura, T. (2017) "Proposal of a Modulated Extended Cumulative Exposure Model for the Step-Up Voltage Test," *In: Ao SI., Kim H., Amouzegar M. (eds) Transactions on Engineering Technologies: World Congress on Engineering and Computer Science 2015, Springer, Singapore*, Chapter 25, Pages 349-363, 査読あり

Hirose, H., Tokunaga, M., Sakumura, T., Sulaiman, J., and Darwis, H. (2015) "Matrix Approach for the Seasonal Infectious Disease Spread Prediction," *IEICE Transactions on Fundamentals*, Volume E98-A, Number 10, Pages 2010-2017, 査読あり

[学会発表](計 12 件)

Kodaira, Y., Abe, K., Sakumura, T. and Kamakura, T., "Bias Reduction of ML Estimators of Gamma Distribution Parameters and Comparison with Other Methods," *2017 Hangzhou International Statistical Symposium*, Hangzhou Huagang HNA Resort, Hangzhou city, Zhejiang province, China, 1-3 November, 2017

作村建紀, 鎌倉稔成, 柳本武美 "指数拡散分布族における母数推定の改良," *2017年度統計関連学会連合大会*, 南山大学, 愛知, 2017年9月

Sakumura, T., "Optimum life test plan for accelerated deterioration test of three factor stress level," *Section on International Statistical Institute, 2017 Joint Statistical Meetings (JSM2017)*, Baltimore, America, 29 July-3 August, 2017

Sakumura, T., "Compromised life test plan for accelerated degradation test of three stress levels," *日本信頼性学会第25回春季信頼性シンポジウム*, 日本科学技術連盟本部, 東京, 2017年5月31日

作村建紀, 鎌倉稔成, 柳本武美, "逆ガウス分布における自然母数の事後平均," *日本計算機統計学会第31回大会*, 東京理科大学, 東京, 2017年5月12日

Yoshida, A., Sakumura, T. and Kamakura, T., "Indoor Location Estimation in NLoS Environment Based on ToA Method with Particle Filter" *2016 In-*

*ternational Conference for JSCS 30th Anniversary in Seattle*, Seattle Central Library, WA, America, 2016.10.16-17

Sakumura, T. and Kamakura, T., "Parameter Estimation for the Modulated Extended Cumulative Exposure Model," *Section on Statistical Computing, 2016 Joint Statistical Meetings (JSM2016)*, Chicago, America, 30 July-6 August, 2016

Sakumura, T. and Hirose, H., "Bias Reduction of Abilities for Adaptive Online IRT Testing Systems," *5th International Conference on Learning Technologies and Learning Environments (LTLE2016)*, Kumamoto, Japan, 10-14 July, 2016

作村建紀, 徳貞美子, 田淵直樹, 廣瀬英雄, "応答アイテム数が少数のときに発生する能力推定値のバイアスについて," *日本計算機統計学会第29回シンポジウム*, まなぼつと幣舞(釧路市生涯学習センター), 北海道, 2015年11月27日-28日

徳貞美子, 作村建紀, 田淵直樹, 廣瀬英雄, "アダプティブ IRT テストにおける最適問題数に関する考察," *日本計算機統計学会第29回シンポジウム*, まなぼつと幣舞(釧路市生涯学習センター), 北海道, 2015年11月27日-28日

Yoshida, A., Sakumura, T. and Kamakura, T., "Indoor Location Estimation Based on ToA Data and Bias Estimation Using Gamma Regression," *2015 International Workshop for JSCS 30th Anniversary in Okinawa*, Okinawa Institute of Science and Technology Graduate University (OIST), Japan, 30-31 October, 2015

Sakumura, T. and Kamakura, T., "Modulated Extended Cumulative Exposure Model with Application to the Step-Up Voltage Test," *World Congress on Engineering and Computer Science 2015 (WCECS 2015)*, San Francisco, USA, 21-23 October, 2015

## 6. 研究組織

### (1) 研究代表者

作村 建紀 (SAKUMURA, Takenori)

中央大学・理工学部・助教

研究者番号: 50735389