

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 13 日現在

機関番号：12612

研究種目：若手研究(B)

研究期間：2015～2016

課題番号：15K16256

研究課題名(和文)信頼性の高いピアアセスメントを実現する階層ベイズ項目反応理論

研究課題名(英文)Hierarchical Bayesian Item Response Theory Model for Reliable Peer Assessment

研究代表者

宇都 雅輝(Uto, Masaki)

電気通信大学・大学院情報理工学研究科・助教

研究者番号：10732571

交付決定額(研究期間全体)：(直接経費) 2,100,000円

研究成果の概要(和文)：本研究では、ピアアセスメントにおける評価の信頼性を改善するために、評価者のバイアスを取り除いて学習者の真の能力を推定できる項目反応モデルを開発した。提案モデルは、多値型項目反応モデルに、信頼性改善に有効な必要最小限の評価者パラメータのみを付与したモデルである。提案モデルは、評価者数の増加に伴うパラメータ数の増加が既存モデルより緩慢となるため、評価者数が増加する場面において、既存モデルよりパラメータ数に対するデータ数が増加し、より高精度なパラメータ推定と能力推定が可能となる。さらに、本研究ではスパースデータからの推定精度を改善するために、階層ベイズモデルを用いたベイズ推定法を提案した。

研究成果の概要(英文)：In peer assessment, a problem remains that reliability depends on the rater characteristics. For this reason, some item response models that incorporate rater parameters have been proposed. Those models are expected to improve the reliability if the model parameters can be estimated accurately. However, when applying them to actual peer assessment, the parameter estimation accuracy would be reduced for the following reasons. 1) The number of rater parameters increases rapidly because the models include higher-dimensional rater parameters. 2) The parameter estimation accuracy from sparse assessment data depends on hand-tuning parameters, called hyperparameters. To solve the problems, we propose a new item response model for peer assessment that incorporates rater parameters to maintain as few rater parameters as possible. Furthermore, this article proposes a parameter estimation method using a hierarchical Bayes model for the proposed model that can learn the hyperparameters from data.

研究分野：教育工学

キーワード：ピアアセスメント 項目反応理論 eラーニング eテストング 教育評価

1. 研究開始当初の背景

近年、社会構成主義的学習観の浸透に伴い、評価を学習と同一の文脈で行う「真正な評価」が注目されている。真正な評価では、現実的で実践的な課題に学習者が取り組み、そのパフォーマンスを学習場面で評価することを重視する。

真正な評価を実現する手法の一つとして、学習者同士の相互評価を表すピアアセスメントが注目されている。ピアアセスメントには、学習者の内省促進、学習意欲の向上、深い学習の促進など多くの利点が報告されており、すでに多くの学習場面で利用されている。

ピアアセスメントは、客観式テストと異なり、採点結果に評価者の主観が反映される点の特徴である。このため、ピアアセスメントでは、評価の信頼性が評価者の特性に依存することが問題とされてきた。

この問題を解決するアプローチのひとつとして、評価者特性を表すパラメータを付与した項目反応理論が提案されてきた。これらの既存の項目反応モデルは、論述式テストの採点のように数名の評価者が多数の受験者を評価するような場面への適用を想定し、比較的高次元の評価者パラメータを採用している点の特徴である。しかし、高次元の評価者パラメータを採用したことにより、既存モデルをピアアセスメントのように評価者数が増加する場合に適用すると、評価者パラメータ数が急速に増加し、パラメータ数に対するデータ数が少なくなってしまう。一般に、パラメータ数に対するデータ数が減少すると、パラメータの推定精度が低下する。

また、既存モデルのパラメータ推定法にはベイズ推定が採用されている。しかし、ベイズ推定は、スパースなデータに適用したとき、推定精度がハイパーパラメータと呼ばれるハンド・チューニングパラメータに依存する問題が残る。ピアアセスメントでは、学習者数が増加すると、全ての学習者間で相互評価を行うことは現実的でなくなる。このような場合、グループ分けなどを行い、一部の学習者間でのみ評価しあうことが一般的といえるが、このときの評価データは非常にスパースになる。

以上の2つの要因により、パラメータの推定精度が低下すると、項目反応モデルを用いても信頼性の高い学習者の能力推定は期待できない。すなわち、既存モデルを、評価者数が増加するピアアセスメントに適用すると、能力測定精度が低下する点が問題となる。

2. 研究の目的

上記の問題を解決するために、本研究では、できる限り評価者パラメータ数が少なくなるように評価者特性パラメータを付与した項目反応モデルを開発する。さらに、パラメータ

推定精度が、ハイパーパラメータに依存する問題を解決するために、提案モデルのためのパラメータ推定法として、階層ベイズモデリングを用いたベイズ推定法を提案した。この推定法では、パラメータ推定過程でハイパーパラメータもデータから最適化できる。

3. 研究の方法

3.1. ピアアセスメントデータと項目反応理論

本研究で想定するピアアセスメントデータ X は、課題 $i \in \{1 \dots I\}$ に対する学習者 $j \in \{1 \dots J\}$ の成果物に、評価者 $r \in \{1 \dots R\}$ が与える評点 $k \in \{1 \dots K\}$ の集合として、以下のような3相データと定義する。

$$X = \{x_{ijr} \mid x_{ijr} \in \{-1, 1 \dots K\}\} \\ (i = 1 \dots I, j = 1 \dots J, r = 1 \dots R) \quad (1)$$

ここで、 $x_{ijr} = -1$ は欠測データを表す。

本研究では、このデータに項目反応理論を適用するが、一般的な項目反応モデルはこのような3相データに対して直接に適用することはできない。この問題を解決するアプローチとして、評価者特性を表すパラメータを加えた項目反応モデルが提案されてきた [1, 2, 3, 4, 5]。しかし、上述の通り、既存モデルでは比較的高次元の評価者パラメータが採用されているため、評価者数が増加するとパラメータ数が増加する。一般に、データ数に対するパラメータ数が増加すると、パラメータの推定精度が低下し、それは学習者の能力推定精度の低下を引き起こす。そこで、本研究では、必要最小限の評価者パラメータのみを採用した新たな項目反応モデルを開発する。

3.2. 提案モデル

提案モデルは、多値型項目反応モデルとして一般的な段階反応モデルに対して、評価者の厳しさと一貫性のパラメータを加えたモデルとして次のように定式化した。

$$P_{ijrk} = P_{ijrk-1}^* - P_{ijrk}^* \quad (2)$$

$$P_{ijrk}^* = \frac{\exp(\alpha_i \alpha_r (\theta_j - b_{ik} - \epsilon_r))}{1 + \exp(\alpha_i \alpha_r (\theta_j - b_{ik} - \epsilon_r))} \quad (3)$$

$$: k = 1, \dots, K - 1$$

ここで、 θ_j は学習者 j の能力、 α_i は課題 i の識別力、 α_r は評価者 r の評価の一貫性、 ϵ_r は評価者 r の評価の厳しさ、 b_{ik} は課題 i において k より大きい評点を得る困難度をそれぞれ表す。また、 $P_{ijr0}^* = 1, P_{ijrK}^* = 0$ とする。このモデルでは、課題の困難度パラメータ b_{ik} に順序制約 $b_{i1} < b_{i2} < \dots < b_{iK-1}$ を課す。モデルの識別性のために、 $\alpha_{r=1} = 1, \epsilon_{r=1} = 0$ を仮定する。

3.3. 提案モデルの特徴

表 1 に、提案モデルと Patz & Junker[1] (Patz1999 と呼ぶ)、宇佐美[2] (Usami2010 と呼ぶ)、Ueno & Okamoto[3] (Ueno2008 と呼ぶ)、Patz, Junker & Johnson[4] (HRM-Patz と呼ぶ)、DeCarlo, Kim & Johnson[5] (HRM-DeCarlo と呼ぶ) のモデルにおけるパラメータ数を示した。

表 1. 各モデルのパラメータ数

	パラメータ数
提案モデル	$IK+2(R-1)+J$
Patz1999	$I(K+R-1)+J$
Usami2010	$IK+3(R-1)+J$
Ueno2008	$2I+R(K-1)-1+J$
HRM-Patz	$I(K-1+J)+2R+J$
HRM-DeCarlo	$I(K+J)+RK+J$

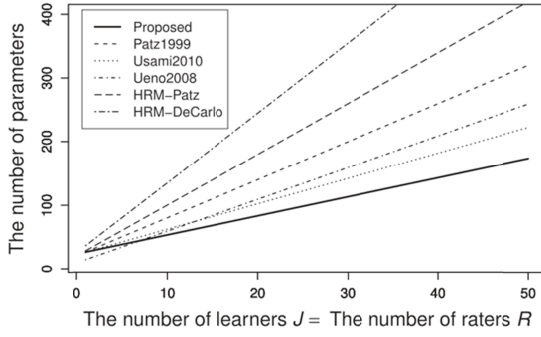


図 1. モデルのパラメータ数

表 1 より、課題数に対して評価者数が多い場合、具体的には、評点数 $K = 5$ としたとき、 $2(R+1) > 3I \wedge I \geq 2$ の場合に提案モデルのパラメータ数が最小となることがわかる。提案モデルは、他の多くのモデルで共通して採用されている重要な課題パラメータと評価者パラメータを残しつつ、評価者パラメータ数を軽減している。そのため、評価者数が多くなるピアアセスメントにおいて、評価者の重要な特性は考慮しつつ、高精度にパラメータと能力を推定できる。

3.4. パラメータ推定

項目反応理論のパラメータ推定法としては、事後密度の期待値を点推定値とする Expected a Posterior (EAP) 推定によるベイズ推定法が一般に頑健であることが知られている。

ベイズ推定では、モデルパラメータを確率変数とみなし事前分布を仮定する。事前分布のパラメータはハイパーパラメータと呼ばれ、分析者の事前知識を反映して主観的に決定される。

ここで、提案モデルパラメータをそれぞれ $\theta = \{\theta_1 \dots \theta_J\}$, $\alpha_i = \{\log \alpha_{i=1}, \dots, \log \alpha_{i=I}\}$, $\mathbf{b} = \{b_{11} \dots b_{IK-1}\}$, $\alpha_r = \{\log \alpha_{r=1} \dots \log \alpha_{r=R}\}$,

$\epsilon = \{\epsilon_1 \dots \epsilon_R\}$ とし、各パラメータへの事前分布を $g(\theta|\tau_\theta)$, $g(\alpha_i|\tau_{\alpha_r})$, $g(\mathbf{b}|\tau_b)$, $g(\alpha_r|\tau_{\alpha_i})$, $g(\epsilon|\tau_\epsilon)$ とする。 τ は、事前分布 g のパラメータ、すなわちハイパーパラメータを表す。このとき、提案モデルの事後確率は次式で表される。

$$g(\theta, \alpha_i, \mathbf{b}, \alpha_r, \epsilon|U) \propto L(U|\theta, \alpha_i, \mathbf{b}, \alpha_r, \epsilon) g(\theta|\tau_\theta) g(\alpha_i|\tau_{\alpha_r}) g(\mathbf{b}|\tau_b) g(\alpha_r|\tau_{\alpha_i}) g(\epsilon|\tau_\epsilon) \quad (4)$$

ただし、

$$L(U|\theta, \alpha_i, \mathbf{b}, \alpha_r, \epsilon) = \prod_{j=1}^J \prod_{i=1}^I \prod_{r=1}^R \prod_{k=1}^K (P_{ijrk})^{z_{ijkr}} \quad (5)$$

$$z_{ijkr} = \begin{cases} 1: x_{ijr} = k \\ 0: otherwise. \end{cases} \quad (6)$$

事前分布として、 $\theta_j, \log \alpha_i, \log \alpha_r, \epsilon_r$ には正規分布を、 $\mathbf{b}_i = \{b_{i1} \dots b_{iK-1}\}$ には多次元正規分布を仮定することが一般的である。

EAP 推定では、事後分布 $g(\theta, \alpha_i, \mathbf{b}, \alpha_r, \epsilon|U)$ の最大化あるいは期待値を求めることでパラメータの点推定値を求める。期待値計算には MCMC を利用することが一般的である。

3.5. 階層ベイズ推定

ベイズ推定は、一般に MMLE と比べて頑健な推定が可能であるが、その推定精度はハイパーパラメータの選択に依存することが知られている。特に、スパースデータに対して不適切なハイパーパラメータを選択するとパラメータの推定精度が著しく低下することが知られている。

この問題を解決するアプローチの一つとして、ハイパーパラメータに対してさらに事前分布を仮定した階層ベイズモデルに基づく推定が提案されている。階層ベイズを用いた推定は機械学習分野で広く利用されてきた手法である。

提案モデルの階層ベイズ事後分布は次式で定義できる。

$$g(\theta, \alpha_i, \mathbf{b}, \alpha_r, \epsilon, \tau_{\alpha_r}, \tau_b, \tau_{\alpha_i}, \tau_\epsilon|U) \propto L(U|\theta, \alpha_i, \mathbf{b}, \alpha_r, \epsilon) g(\theta|\tau_\theta) g(\alpha_i|\tau_{\alpha_r}) g(\tau_{\alpha_r}) g(\mathbf{b}|\tau_b) g(\tau_b) g(\alpha_r|\tau_{\alpha_i}) g(\tau_{\alpha_i}) g(\epsilon|\tau_\epsilon) g(\tau_\epsilon) \quad (7)$$

本研究では、式(7)に基づく期待事後確率推定値 (EAP) を求める手法として、マルコフ連鎖モンテカルロ法 (Markov Chain Monte Carlo: MCMC) に基づく推定法を定式化した。詳細は、Uto and Ueno (2016) を参照されたい。

4. 研究成果

本研究では、シミュレーション実験により

既存モデルと提案モデルの比較を行い、提案モデルが既存モデルより高精度なパラメータ推定と能力推定を実現できることを示した。また、階層ベイズ推定を用いることで、その性能をさらに向上できることが明らかとなった。

また、被験者実験により実際のピアアセスメントデータを収集し、それをを用いて提案モデルの有効性を評価した。実験の結果、提案モデルが実データに最もよく当てはまるモデルであることが確認できた。また、能力評価の信頼性の観点でも既存モデルを上回る性能を示すことが明らかとなった。シミュレーション実験同様、階層ベイズ推定を用いることで、提案モデルの性能をさらに向上できることが明らかとなった。

提案モデルに関する研究成果は、電子情報通信学会の論文誌(宇都・植野, 2015)に掲載された。さらに、提案モデルのパラメータ推定法に階層ベイズを用いた拡張は、分野のトップジャーナルである IEEE Transactions on Learning Technologies (Uto and Ueno, 2016)に掲載されている。また、提案モデルを含む、評価者パラメータを付与した項目反応モデルについて、様々な観点で比較・考察したレビュー論文が、テスト分野の専門論文誌である日本テスト学会論文誌(宇都・植野, 2016)に掲載されている。

さらに、本研究の途中成果は、教育工学分野のトップ国際会議である AIED を含め、国内外の多くの学会で発表した。また、提案モデルを、ピアアセスメントのためのグループ構成最適化や、ルーブリック評価などに応用する研究も進行しており、現在までの成果は主に国内学会においてすでに発表している。

また、本研究の成果は、社会的にも高いインパクトを有しており、特に入試分野においてはその実用化が強く期待されている。2020年の入試改革により、大学入試に論述式・記述式試験の導入が検討されており、それに伴う評価者評価の信頼性を向上するために提案手法が有効であると解釈できる。すでにリクルートキャリアやベネッセホールディングス、日本英語検定協会、医学系大学共用試験などにおける、グループディスカッションや論述式試験、スピーキング試験や実技試験などに提案手法を適用する実践も開始しており、本研究の成果はますます波及すると期待できる。

<参考文献>

- [1] Patz, R. J. & Junker, B. W. (1999b), Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. Journal of Educational and Behavioral Statistics, 24, 342-366.
- [2] 宇佐美慧 (2010). 採点者側と受験者側のバイアス要因の影響を同時に評価す

る多値型項目反応モデル:MCMC アルゴリズムに基づく推定, 教育心理学研究, 58, (2), 163-175.

- [3] Ueno, M. & Okamoto, T. (2008). Item response theory for peer assessment, IEEE International Conference on Advanced Learning Technologies, 554-558
- [4] Patz, R. J., Junker, B. W. & Johnson, M. S. (1999a). The hierarchical rater model for rated test items and its application to large-scale educational assessment data, Journal of Educational and Behavioral Statistics, 27, (4), 341-366.
- [5] DeCarlo, L. T., Kim, Y. K. & Johnson, M. S. (2011). A hierarchical rater model for constructed responses, with a signal detection rater model. Journal of Educational Measurement, 48, (3), 333-356.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 3 件)

Masaki Uto, Maomi Ueno (2016) Item Response Theory for Peer Assessment. IEEE Transactions on Learning Technologies, Vol.9, No.2, pp.157-170【査読有】

<http://doi.ieeecomputersociety.org/10.1109/TLT.2015.2476806>

宇都雅輝, 植野真臣 (2016) パフォーマンス評価のため項目反応モデルの比較と展望. 日本テスト学会誌, Vol.12, No.1, pp.55-75【査読有】

<http://www.jartest.jp/book12-1.html>

宇都雅輝, 植野真臣 (2015) ピアアセスメントの低次評価者母数をもつ項目反応理論. 電子情報通信学会論文誌 D, Vol. 98, No. 1, pp. 3-16【査読有】
http://search.ieice.org/bin/summary.php?id=j98-d_1_3

[学会発表](計 13 件)

Masaki Uto, Nguyen Duc Thien, Maomi Ueno. Group optimization to maximize peer assessment accuracy using item response theory. International Conference on Artificial Intelligence in Education(国際会議), 2017年6月28日~2017年7月2日, Wuhan (China).

Taiyo Utsuhara, Masaki Uto, Asana Ishihara, Koichi Ota, Ayako Hirano, Atsushi Yoshikawa, Maomi Ueno.

Features of Globalization in Japanese Graduate Schools. International Conference on Education (国際会議), 2017年3月12日～2017年3月15日, San Diego (United States of America).
木下涼, 宇都雅輝, 植野真臣. 足場かけに基づくアカデミックライティング学習支援システム. 教育システム情報学会学生研究発表会. 2017年2月28日～2017年3月1日, 銚子市民センター (千葉県銚子市).
宇都雅輝, 植野真臣. パフォーマンス評価の多様なバイアスを考慮した項目反応モデル. 日本教育工学会第32回全国大会, 2016年9月17日～2016年9月19日, 大阪大学 (大阪府吹田市)
Thien Nguyen, Masaki Uto, Maomi Ueno. Group formation for peer assessment using item response theory. 日本教育工学会第32回全国大会, International Session, 2016年9月17日～2016年9月19日, 大阪大学 (大阪府吹田市)
宇都雅輝, 植野真臣. 評価者と課題の多様な特性を考慮した項目反応モデル. 日本テスト学会14回全国大会, 2016年9月8日～2016年9月9日, 電気通信大学 (東京都調布市)
Thien Nguyen, 宇都雅輝, 植野真臣. ピアアセスメントの精度を最適化する評価者選択手法. 日本テスト学会14回全国大会, 2016年9月8日～2016年9月9日, 電気通信大学 (東京都調布市)
山本美紀, 宇都雅輝, 西山悠, 川野秀一, 植野真臣. Co-creating ルーブリックの自己制御学習および自己評価力への影響分析. 日本テスト学会14回全国大会, 2016年9月8日～2016年9月9日, 電気通信大学 (東京都調布市)
Nguyen Duc Thien, Masaki Uto, Yu Abe, Maomi Ueno. Reliable Peer Assessment for Team-project-based Learning using Item Response Theory. International Conference on Computers in Education (国際会議), 2015年11月30日～2015年12月4日, Hangzhou (China).
山本美紀, 宇都雅輝, 植野真臣. 項目反応理論によるルーブリックの自己評価力への影響分析-評価者特性と目標志向性, 学習観, 動機づけ, 学習方略に着目して. 日本教育工学会第31回全国大会, 2015年9月21日～2015年9月23日, 電気通信大学 (東京都調布市)
宇都雅輝. ピアアセスメントのための項目反応理論を用いた評価者選択. 教育システム情報学会第40回全国大会, 2015年9月1日～2015年9月3日, 徳島大学 (徳島県徳島市)
Masaki Uto, Maomi Ueno. Academic Writing Support System using Bayesian

Networks. IEEE International Conference on Advanced Learning Technologies (国際会議), 2015年7月6日～2015年7月9日, Hualien (Taiwan).
Masaki Uto, Maomi Ueno. Item Response Model with Lower Order Rater Parameters for Peer Assessment. International Conference on Artificial Intelligence in Education (国際会議), 2015年6月22日～2015年6月26日, Madrid (Spain)

〔その他〕
ホームページ等
<https://sites.google.com/site/utomasaki>

6. 研究組織

(1) 研究代表者

宇都 雅輝 (UTO, Masaki)
電気通信大学・大学院情報理工学研究科・助教
研究者番号: 10732571