ピアアセスメントの低次評価者母数を持つ階層ベイズ項目反応理論

Hierarchical Bayesian Item response theory with lower order rater parameters for peer assessment

宇都雅輝 *1， 植野真臣 *2
Masaki Uto*1, Maomi Ueno*2
*1 長岡技術科学大学, *2 電気通信大学大学院情報システム学研究科
*1Nagaoka University of Technology, *2Graduate School of Information System, University of Electro Communication
Email: uto@oberon.nagaokaut.ac.jp

あらまし： ピアアセスメントにおける課題の一つとして，評価の信頼性が評価者の特性に依存する問題が指摘されている．この問題を解決するために，評価者特性を表す母数を付加した項目反応理論が提案されてきた．しかし，既存モデルでは，ピアアセスメントのように評価者数が増加したとき，評価者母数数に対するデータ数が少なくなり，高精度な母数推定が期待できない．また，先行研究で採用しているベイズ推定法では，母数推定精度が，分析者が決定する事前分布の母数（超母数）に依存する．以上の問題を解決するために，本研究では，通常の項目反応モデルに対し，できる限り評価者母数が少なくなるように，評価者の「評価の一貫性」と「評価の厳しさ」を表す母数を付加した，ピアアセスメントのための新たな項目反応理論を提案する．さらに，超母数に依存しない提案モデルの母数推定法として，超母数に事前分布を仮定する階層ベイズモデルを用いた推定法を提案する．
キーワード： ピアアセスメント，項目反応理論，信頼性，評価者特性，階層ベイズ

1　はじめに

近年，構成主義における学習評価法として，学習者同士による学習成果物の相互評価法，ピアアセスメントが注目されている．ピアアセスメントは多数の利点を持つことが報告されており [1]，これまでに様々な学習場面で利用されてきた．

一方，ピアアセスメントの課題として，評価の信頼性が評価者の特性に依存する問題が指摘されている [1][2]．具体的には，次のような評価者特性が評価の信頼性低下を引き起こす．1) 評価者間で評価の甘さ/厳しさが存在すること．2) 評価者間/評価者内で評価基準が一貫している保証がないこと．

同様の課題は，複数の評価者による論述式テストの採点などでも指摘されており，この問題を解決するために，評価者特性を表す母数を付加した項目反応理論が提案されてきた（例えば，[1][2][3][4]）．

しかし，既存モデルでは，ピアアセスメントのように評価者数が増加したとき，評価者母数数に対するデータ数が少なくなり，高精度な母数推定が期待できない．また，先行研究で採用しているベイズ推定法では，母数推定精度が，分析者が決定する事前分布の母数（超母数）に依存する [5]．

以上の問題を解決するために，本研究では，通常の項目反応モデルに対し，できる限り評価者母数が少なくなるように，評価者の「評価の一貫性」と「評価の厳しさ」を表す母数を付加した，ピアアセスメントのための新たな項目反応理論を提案する．さらに，超母数に依存しない提案モデルの母数推定法として，超母数に事前分布を仮定する階層ベイズモデルを用いた推定法を提案する．提案手法の特徴は次の通りである．1) 既存手法よりも高精度な母数推定が可能である．2) 評価者の評価の一貫性と厳しさの影響を反映した学習者の能力推定が可能である．3) 評価の信頼性向上が期待できる．

2　評価者特性を考慮した項目反応理論

本研究では，学習者 $j\,(j=1,\cdots,J)$ の課題 $i\,(i=1,\cdots,I)$ に対する評価者 r（$r=1,\cdots,R$）の評価カテゴリ k（$k=1,...,K$）で構成される「学習者」×「課題」×「評価者」の三相データ $\boldsymbol{U}$ を扱う．通常の項目反応モデルは，このような三層データに直接には適用できない．この問題を解決するために，多値型項目反応モデルとして知られる一般化部分採点モデルや段階反応モデルに対し，評価者特性を表す母数を付与した項目反応モデルが提案されてきた [1][2][3][4].

例えば，宇佐美 [2] は，一般化部分採点モデルに評価者パラメータを加えた以下のモデルを提案している．

$$P_{ijrk}=\frac{\exp\sum_{m=1}^{k}\left[\alpha_i\alpha_r(\theta_j-(\beta_i+\beta_r)-d_{im}d_r)\right]}{\sum_{l=1}^{K}\exp\sum_{m=1}^{l}\left[\alpha_i\alpha_r(\theta_j-(\beta_i+\beta_r)-d_{im}d_r)\right]}$$

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，β_i は課題 i の位置母数，β_r は評価者 r の位置母数，d_{ik} は課題 i における評点 k の閾値母数（ただし $d_{i1}=0$），d_r は評価者 r の閾値母数を表す．母数の識別性のために，$\Pi_r\alpha_r=1$，$\sum_r\beta_r=0$，$\Pi_r d_r=1$ を仮定する．

植野ら [1] は，ピアアセスメントのための項目反応理論として，段階反応モデルを拡張した以下のモデルを提案している．

$$P_{ijrk}=P^*_{ijrk-1}-P^*_{ijrk}$$
$$P^*_{ijrk}=[1+\exp(-\alpha_i(\theta_j-b_i-\varepsilon_{r,k}))]^{-1}$$

ただし，$P^*_{ijr0}=1$, $P^*_{ijrK}=0$ とする．ここで，b_i は課題 i の難易度，$\varepsilon_{r,k}$ は評価者 r による評点 k への厳しさを表す（ただし $\varepsilon_{r,1}<\varepsilon_{r,2}<\cdots<\varepsilon_{r,K-1}$）．

しかし，これらのモデルは，評価者数が増加したとき，評価者母数数に対するデータ数が少なくなり，高精度な母数推定が期待できない．

また，先行研究の多くでは，母数推定法としてベイズ推定法を採用している．具体的には，各母数に事前分布を仮定して事後分布を導出し，マルコフ連鎖モンテカルロ (MCMC) 法を用いて，各母数の EAP(Expected a posterior) 推定値を求めている．しかし，この推定法では，分析者が決定する超母数に母数推定精度が依存することが知られており [5]，先行研究のような恣意的な超母数の決定は好ましくない．

以上の問題を解決するために，本研究では，通常の項目反応モデルに対し，できる限り評価者母数が少なくなるように，評価者の「評価の一貫性」と「評価の厳しさ」を表す母数を付加したモデルを提案する．さらに，超母数に依存しない提案モデルの母数推定法として，超母数に事前分布を仮定する階

層ベイズモデルを用いた推定法を示す．

3 提案モデル

本研究では，ピアアセスメントにおける項目反応理論として，段階反応モデルを拡張した以下のモデルを提案する．

$$P_{ijrk} = P^*_{ijrk-1} - P^*_{ijrk}$$
$$P^*_{ijrk} = [1 + \exp(-\alpha_i \alpha_r (\theta_j - b_{ik} - \varepsilon_r))]^{-1}$$

ただし，$P^*_{ijr0} = 1$，$P^*_{ijrK} = 0$ とする．ここで，b_{ik} は課題 i において評点 k を得る難易度，ε_r は評価者 r の評価の厳しさを表す．ただし $b_{i1} < b_{i2} < \cdots < b_{iK-1}$ とする．また，母数の識別性のために，$\Pi_r \alpha_r = 1$ を仮定する．

提案モデルでは，評価者の「評価の一貫性」と「評価の厳しさ」母数を，評価者一人につき一つずつになるように付与している．これにより，評価者数の増加に伴う評価者母数数の増加が既存モデルに比べて緩慢となる．

提案モデルの母数数は，$K = 5$ とすると，$2R > 3I$ かつ $I > 2$ のとき，既存モデル [1][2][3] と比べて最も少なくなる．これらの条件は，課題数に対し評価者数が多いピアアセスメントでは一般に満たされるといえる．

図 1 に，評価者母数の異なる 2 名の評価者における，提案モデルの反応曲線を示す．図 1 では，横軸が θ，縦軸が各評点が付与される確率を表す．評価者母数は，Rater1(左図) は $\alpha_r = 1.5, \varepsilon_r = 1.0$，Rater2(右図) は $\alpha_r = 0.8, \varepsilon_r = -1.0$ とした．

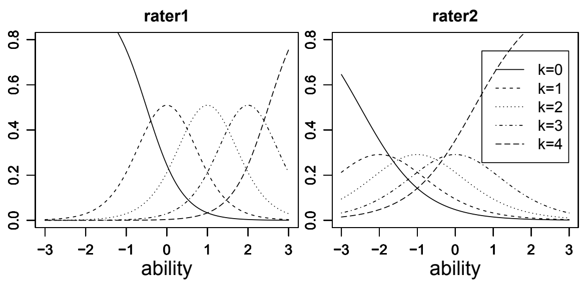

図 1 異なる評価者による提案モデルの反応関数

図 1 より，評価の一貫性が高い Rater1 の分布では，θ が変化したときの各評点への反応確率の変化が大きいことがわかる．これは，Rater1 の方が学習者の能力を精度良く識別できることを意味する．また，Rater1 は，評価の厳しさ ε_r が大きく，反応曲線は全体として右に移動していることがわかる．これは，Rater1 から高い評点を得るには，Rater2 から同じ評点を得るより高い能力が必要であることを意味する．

次に，提案モデルの階層ベイズ事後分布を導出する．ここで，各母数の集合を $\boldsymbol{\theta} = \{\theta_j | j = 1, \cdots, J\}$，$\boldsymbol{\alpha_i} = \{\alpha_i | i = 1, \cdots, I\}$，$\boldsymbol{b} = \{\boldsymbol{b_i} | i = 1, \cdots, I\}$，$\boldsymbol{b_i} = \{b_{ik} | k = 1, \cdots, K-1\}$，$\boldsymbol{\alpha_r} = \{\alpha_r | r = 1, \cdots, R\}$，$\boldsymbol{\varepsilon} = \{\varepsilon_r | r = 1, \cdots, R\}$，各母数の超母数を $\tau_\theta, \tau_{\alpha_i}, \tau_b, \tau_{\alpha_r}, \tau_\varepsilon$，尤度を $L(\boldsymbol{U}|\boldsymbol{\theta}, \boldsymbol{\alpha_i}, \boldsymbol{b}, \boldsymbol{\alpha_r}, \boldsymbol{\varepsilon})$ で表す．このとき，データ $\boldsymbol{U}$ を所与として，提案モデルの階層ベイズ事後分布は以下のように書ける．

$$\begin{aligned} &g(\boldsymbol{\theta}, \tau_\theta, \boldsymbol{\alpha_i}, \tau_{\alpha_i}, \boldsymbol{b}, \tau_b, \boldsymbol{\alpha_r}, \tau_{\alpha_r}, \boldsymbol{\varepsilon}, \tau_\varepsilon | \boldsymbol{U}) \\ &\quad \propto L(\boldsymbol{U}|\boldsymbol{\theta}, \boldsymbol{\alpha_i}, \boldsymbol{b}, \boldsymbol{\alpha_r}, \boldsymbol{\varepsilon}) g(\boldsymbol{\theta}|\tau_\theta) g(\tau_\theta) g(\boldsymbol{\alpha_i}|\tau_{\alpha_i}) g(\tau_{\alpha_i}) \\ &\qquad g(\boldsymbol{b}|\tau_b) g(\tau_b) g(\boldsymbol{\alpha_r}|\tau_{\alpha_r}) g(\tau_{\alpha_r}) g(\boldsymbol{\varepsilon}|\tau_\varepsilon) g(\tau_\varepsilon) \quad (1) \end{aligned}$$

ここでは，θ_j,$\log \alpha_i$,$\log \alpha_r$, ε_r の事前分布に正規分布を用いる．階層ベイズでは，これらの正規分布の母数（平均 μ_τ，分散 σ^2_τ）に対してさらに事前分布（超事前分布と呼ぶ）を仮定する．超事前分布は共益事前分布とし，μ_τ の事前分布には正規分布，分散 σ^2_τ には逆ガンマ関数を用いる．他方，$\boldsymbol{b_i}$ の事前分布には，平均値ベクトル $\boldsymbol{\mu}_b$，共分散行列 $\boldsymbol{\Sigma}_b$ の $K-1$ 次元正規分布を用いる．超事前分布として，$\boldsymbol{\mu}_b$ の事前分布には $K-1$ 次元正規分布，$\boldsymbol{\Sigma}_b$ には逆ウィシャート分布を用いる．

ここでは，式 (1) の階層ベイズ事後分布に対し，MCMC 法を適用し，提案モデルの母数と超母数を同時に推定する．MCMC 法には，Patz et al.[3] や宇佐美 [2] でも採用している Metropolis hastings と Gibbs sampling を組み合わせた手法を用いる．階層ベイズモデルでは，通常のベイズモデルの場合と異なり，MCMC の各更新ステップに超母数の更新が付与される．母数の EAP 推定値は，MCMC から得られたサンプルの平均値として求められる．

4 実データへの適用

本研究では，提案モデルの妥当性を確認するために，次の被験者実験を行った．

1) 著者らの一人が開講している統計学の e-learning 講義から 5 課題× 20 名分のレポートを収集し，本講義の受講経験のある理系大学院生 20 名に採点させた．採点は，筆者らが用意したルーブリックを用いて 5 段階で行わせた．2) 収集した評価データを用いて，提案モデル，宇佐美 [2]，植野ら [1] の母数を MCMC で推定し，モデル選択基準の一つとして知られる BIC を算出した．BIC では値が最大となるモデルを最適モデルとみなす．3) J=R=5, 10 になるように抽出したデータについても同様に BIC を求めた．

実験結果を表 1 に示す．表 1 から，全ての場合で提案モデルが最適なモデルとして推定されたことがわかる．

表 1 実データによるモデル比較

	J=R=5	J=R=10	J=R=20
提案モデル	-209.55	-394.85	-1511.71
宇佐美 [2]	-239.48	-443.30	-1593.89
植野ら [1]	-222.68	-472.75	-1734.20

5 おわりに

本稿では紙面の都合上割愛したが，母数推定精度や信頼性に関する実験でも，提案手法が優れた性能を示すことを確認している．発表ではこれらの実験結果の報告も予定している．

参考文献

[1] 植野真臣，ソンムァンポクポン，岡本敏雄，永岡慶三，“ピアアセスメントにおける評価者特性を考慮した項目反応理論，” 電子情報通信学会論文誌. D, 情報・システム，vol.91，no.2，pp.377–388，2008.

[2] 宇佐美慧，“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，vol.58，no.2，pp.163–175，2010.

[3] R.J. Patz and B.W. Junker, “Applications and extensions of mcmc in irt: Multiple item types, missing data, and rated responses,” Journal of Educational and Behavioral Statistics, vol.24, pp.342–366, 1999.

[4] J.M. Linacre, Many-faceted Rasch Measurement, MESA Press, 1989.

[5] J.P. Fox, Bayesian Item Response Modeling: Theory and Applications, Springer, 2010.