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
Rater Selection for Peer Assessment using Item Response Theory

宇都雅輝 *1
Masaki Uto*1
*1 長岡技術科学大学
*1Nagaoka University of Technology
Email: uto@oberon.nagaokaut.ac.jp

あらまし： 大規模なピアアセスメントでは，一人の学習者に数名の評価者を割り当てて評価を行うことが多い．しかし，ピアアセスメントでは，能力評価の精度が評価者の特性に依存することが知られており，学習者ごとに異なる評価者集団が評価を行った場合，学習者ごとに能力推定精度が異なる問題が生じる．この問題を解決するために，本研究では，評価者パラメータを加えた項目反応理論を用いて，全ての学習者に同程度の能力推定精度を与えるような評価者割り当て手法を提案する．

キーワード： ピアアセスメント，評価者選択，項目反応理論，等質テスト構成

1 はじめに

近年，Massively Open Online Courses (MOOCs) に代表される大規模 e ラーニングにおいて，学習者同士の相互評価法，ピアアセスメントが注目されている [1]．学習場面におけるピアアセスメントは，これまで，学習者同士で形成的フィードバックを行わせることによる学習支援を目的として利用されることが一般的であった．しかし，近年では，レポートなどのパフォーマンス評価にピアアセスメントを利用する研究や実践も広まりつつある．ピアアセスメントによるパフォーマンス評価は，学習者数が大幅に増加しても教員の負荷を増やすことなく評価を実施できる利点がある [2]．

一方で，ピアアセスメントの課題の一つとして，評価の信頼性が評価者の特性に依存する問題が知られている [1, 2, 3]．この問題を解決するために，評価者特性パラメータを加えた項目反応理論が提案されてきた [2, 3]．これらの項目反応理論では，評価者特性を考慮して学習者の能力を推定するため，平均点などの素点に基づく評価法より精度良く学習者の能力を推定できる [3]．

しかし，大規模なピアアセスメントでは，学習者ひとりに対して数名の評価者を割り当てて評価を行うことが一般的であり，このようなピアアセスメントにこれらの項目反応モデルを適用した場合，以下の問題が生じる．

1. 各学習者の能力推定精度は，割り当てられた評価者の特性に依存するため，学習者間で能力推定精度に差違が生じる．
2. 学習者ひとりあたりの評価者数が減少するため，能力推定精度が低下する．

これらの問題を解決するために，本研究では，学習者ひとりに対して数名の評価者を割り当てるピアアセスメントにおける評価者割り当て手法を提案する．本研究では，各学習者に割り当てられる評価者集団が，能力推定精度に関して等質な特性を持ちつつ，それらの精度が出来る限り大きくなるような割り当てを目指す．

本研究に類似する問題として，テスト理論における等質テスト構成が知られている [4]．等質テスト構成は，アイテムバンク内のテスト項目を組み合わせて，受験者の能力推定精度に関する特性が等質となる複数テストフォームを構成する問題である．等質テスト構成では，一般に，項目反応理論における能力推定精度の指標のひとつであるフィッシャー情報量が等質になるように複数テストを構成する．これらの問題は，整数計画問題に帰着させて定式化されることが一般的である (例えば [4])．

本研究の主なアイデアは，ピアアセスメントにおける評価者割り当て問題を，テスト項目を評価者，テストフォームを評価者集団とみなした等質テスト構成のメタファと捉え，評価者特性パラメータを加えた項目反応モデルを用いて，情報量が等質となる評価者集団を複数構成し．各学習者に割り当てることにある．しかし，本研究における評価者割り当て問題は，等質テスト構成と以下の点で異なる．

1. 等質テスト構成では，一般に，テスト間でのテスト情報量の差異を最小化することを目指す．これに対し，本研究では，各評価者集団の情報量が，所望の等質条件を満たしつつ最大になるように，各学習者に対する評価者集団を構成することを目指す．
2. ピアアセスメントでは，一人の評価者が評価できる学習者数に限りがあるため，評価者ひとりに対する学習者数に上限を与える必要がある．

そこで，本研究では，ピアアセスメントにおける評価者割り当て問題を，等質テスト同時構成法 [4] を拡張した整数計画問題として定式化する．具体的には，全ての評価者集団の情報量が一定の誤差に収まる条件のもとで，それらの情報量の下限を最大化する Maxmin 問題として定式化する．さらに，提案手法には，制約条件として，評価者ひとりに対する学習者数の上限を与える．

また，本論では，シミュレーション実験により提案手法の有効性を評価する．

2 ピアアセスメントにおける項目反応理論

本研究では，学習者 $j \in \{1, \cdots, J\}$ の課題 $i \in \{1, \cdots, I\}$ に対する評価者 $r \in \{1, \cdots, R\}$ の評価カテゴリ $k \in \{1, ..., K\}$ で構成されるピアアセスメントデータ $\boldsymbol{U} = \{x_{ijr} | x_{ijr} \in \{1, \cdots, K\}, \forall i, \forall j, \forall r\}$ を想定する．また，z_{ijr} を，課題 i において評価者 r が学習者 j に割り当てられるとき 1，そうでないとき 0 を取る変数とするとき，評価者割り当てを $\boldsymbol{Z} = \{z_{ijr} | \forall i, \forall j, \forall r\}$ で表す．

本研究では，このようなピアアセスメントデータに適用できる項目反応モデルとして，評価者数が増加しても高精度な能力推定が期待できる宇都・植野 [3] のモデルを採用する．このモデルでは，課題 i に対する学習者 j の学習成果物に，評

価者 r が評点 k を与える確率を次式で表す．

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

ここで，$p^*_{ijr0} = 1$，$p^*_{ijrK} = 0$ とし，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，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 $b_{i1} < \cdots < b_{iK-1}$．

このモデルにおいて，課題 i における評価者 r のフィッシャー情報量は，能力母数 θ_j を所与として次式で与えられる．

$$I_{ir}(\theta_j) = \alpha_i^2 \alpha_r^2 \sum_k \frac{\left(p^*_{ijrk-1} q^*_{ijrk-1} - p^*_{ijrk} q^*_{ijrk}\right)^2}{p_{ijrk}}$$

ここで，$q^*_{ijrk} = 1 - p^*_{ijrk}$．

さらに，課題 i において学習者 j に与えられる情報量は，評価者割り当て $\boldsymbol{Z}$ を所与として，学習者 j に割り当てられた全ての評価者の情報量の和として，$I_i(\theta_j) = \sum_r I_{ir}(\theta_j) \cdot z_{ijr}$ で与えられる．

本研究では，課題 i において，各学習者に割り当てられる評価者集団の情報量が等質になるような評価者割り当てを目指す．類似の問題として，情報量が所望の等質制約を満たす複数テストフォームを構成する等質テスト構成が知られている．本研究では，評価者割り当て問題を，テスト項目を評価者，テストフォームを評価者集団とみなした等質テスト構成のメタファと捉える．これにより，評価者割り当て問題は，等質テスト構成の文脈において，学習者数と同数の等質テストを構成する問題とみなせる．固定数の等質テストを構成する手法としては，Boekkooi-Timminga[4] の等質テスト同時構成が知られている．

3　等質テスト構成

Boekkooi-Timminga[4] は，F 個の等質テスト $\boldsymbol{f}$ を同時に構成する問題を，以下の整数計画問題として定式化している．

minimize: e

subject to: $\int |I_i(\theta) \cdot x_{if} - T(\theta)\ | d\theta \leq e : \forall f \in \boldsymbol{f}$

ここで，x_{if} は，テスト $f \in \boldsymbol{f}$ に項目 i が含まれるとき 1，そうでないとき 0 を取る変数とし，$T(\theta)$ は，能力値 θ における情報量の目標値を表す．

等質テスト同時構成では，構成されたテストの情報量と目標情報量の差異を最小化するように目的関数が定義されている．一方で，本研究では，構成される評価者集団の情報量と目標情報量との差異が所望の誤差以下となる条件のもと，できる限り情報量が大きくなるように複数の評価者集団を構成する．そこで，本研究では，評価者割り当て問題を，等質テスト同時構成を拡張した問題として定式化する．

4　提案手法

本研究では，課題 i における評価者割り当て問題を，以下の整数計画問題として定式化する．

maximize: y

subject to:

$\int |I_i(\theta_j) - y \cdot T(\theta_j)\ | d\theta_j \leq e : \forall j$

$\sum_j z_{ijr}\ \leq n_u : \forall r$

$\sum_r z_{ijr} \geq 1 : \forall j$

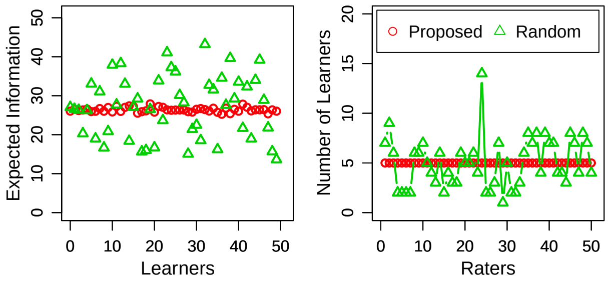

図 1　各学習者に割り当てられた評価者集団の情報量期待値（左図），各評価者が評価する学習者数（右図）

$z_{ijr} = 0 : \forall (j = r)$

提案手法では，構成される評価者集団の情報量関数と目標情報量との差異が e 以下となる条件のもと，目標情報量 $yT(\theta_j)$ を最大化する問題として，評価者割り当て問題を定式化している．さらに，ピアアセスメントでは，ひとりの評価者が評価できる学習者数に限りがあるため，2 つ目の制約式により評価者ひとりに対する学習者数に上限 n_u を与えている．

この整数計画問題を解くことで，等質な情報量特性を保証しつつ，できる限り情報量が大きくなるような評価者割り当て $\boldsymbol{Z}$ が得られる．また，得られた評価者割り当ては，評価者が評価する学習者数の上限が保証されており，特定の評価者への評価負担の集中を回避できる．

5　シミュレーション実験

提案手法の性能を評価するために，シミュレーション実験を行った．ここでは，提案手法とランダムに n_u 人の評価者を割り当てる手法を用いて，次の手順で実験を行った．(1) 課題数 $I = 1$，学習者数 $J = 50$ において，項目反応モデルのパラメータをランダムに生成した．(2) 提案手法とランダム手法を用いて，評価者割り当てを行った．ここで，$n_u = 5$, $e = 1.0$ とした．(3) 得られた評価者割り当て $\boldsymbol{Z}$ について，各学習者に割り当てられた評価者集団の情報量期待値 $\int I_i(\theta_j) d\theta_j$ と，評価者一人あたりの学習数を求めた．

結果を図 1 に示す．図 1 より，提案手法では，各学習者に割り当てられた評価者集団の情報量が等質化されており，評価者一人あたりの学習者数が上限以下となることがわかる．

6　今後の課題

本研究では，評価者・課題パラメータが既知の元で，情報量が等質となるような評価者集団を構成する手法を提案した．今後は，より詳細なシミュレーション実験と被験者実験により提案手法の有効性を確認する．

参考文献

[1] Hoi Suen. Peer assessment for massive open online courses (MOOCs). *The International Review of Research in Open and Distributed Learning*, Vol. 15, No. 3, pp. 313–327, 2014.

[2] M. Ueno and T. Okamoto. Item response theory for peer assessment. In *Advanced Learning Technologies, 2008. ICALT '08. Eighth IEEE International Conference on*, pp. 554–558, 2008.

[3] 宇都雅輝, 植野真臣. ピアアセスメントの低次評価者母数を持つ項目反応理論. 電子情報通信学会論文誌. D, Vol. 98, No. 1, pp. 3–16, 2015.

[4] Ellen Boekkooi-Timminga. *The Construction of Parallel Tests from Irt-Based Item Banks*, Vol. 15. J. Educational Statistics, 1990.