

# 評価者特性の時間変動を考慮した項目反応モデル

1810519 林真由

指導教員 宇都 雅輝 准教授

## 1 はじめに

近年、大学入試や資格試験、教育評価などの場において、パフォーマンス評価は重要な役割を果たしている。その中で、パフォーマンス評価では、評価者の厳しさや一貫性の違い、各得点の使用傾向の差などにより、採点に偏りが生じ、受験者の能力を正確に測ることができないという問題が発生することがある。この問題を解決するために、項目反応理論を使用した能力推定が行われている。最も単純な項目反応モデルは、受験者の能力と課題の困難度の2つのパラメータからなるモデルである。しかし、このモデルは複数の評価者が採点を行うパフォーマンス評価には適用することができない。なぜなら、評価者によって一貫性や厳しさは様々であり、これらを考慮する必要があるためである。このため、これらに評価者特性の考慮を加えた項目反応モデルも多数提案されている。これらのモデルは評価者の基準が採点過程で変化しないという仮定のもと成り立っているが、その仮定は成り立たないことが多い。この評価者の採点基準が採点過程で変化する特性は評価者ドリフト (Rater Drift) と呼ばれ、これを考慮したモデルが提案されている。しかし、既存モデルでは各時間区分ごとのパラメータが独立して推定が難しいという問題点がある。そのため本研究では、時間区分ごとの評価者の厳しさにマルコフ性を仮定した新しい項目反応モデルを提案する。また、シミュレーション実験と実データ実験を通して提案モデルの有効性を示す。

## 2 項目反応理論

本研究は、課題・評価者・時間の特性を考慮した高精度な能力推定を行うことを目的とする。このような能力推定を実現するために、本研究ではIRTを利用する。

IRTは、コンピュータ・テストの普及とともに、近年様々な分野で実用化が進められている数理モデルを用いたテスト理論の一つで、テストを作成・実施・評価・運用するための実践的な数理モデルである。

$$P_{ijrk} = \frac{\exp \sum_{m=1}^k (\theta_j - \beta_i - \beta_r - \pi_r \beta_{rt} - d_m)}{\sum_{l=1}^K \exp \sum_{m=1}^l (\theta_j - \beta_i - \beta_r - \pi_r \beta_{rt} - d_m)}$$

## 3 提案モデル

提案モデルでは、受験者  $j$  のパフォーマンスに、評価者  $r$  が時間区分  $t$  において評点  $k$  を与える確率  $P_{jrk}$  を次式で定義する。今研究では、課題数は1と仮定し、課題パラメータは使用しないこととする。

$$P_{jrk} = \frac{\exp \sum \alpha_r (\theta_j - \beta_{rt} - d_{rk})}{\sum \exp \sum \alpha_r (\theta_j - \beta_{rt} - d_{rk})}$$

$$\beta_{rt} \sim N(\beta_{r(t-1)}, \sigma)$$

$$\beta_{r1} = 0$$

$$\sigma \sim \text{LN}(-3, 0)$$

ここで、 $\alpha_r$  は評価者  $r$  の一貫性、 $\theta_j$  は受験者  $j$  の能力、 $\beta_{rt}$  は評価者  $r$  の時間区分  $t$  における厳しさ、 $d_{rk}$  は評価者  $r$  からスコア  $k$  を得る困難度を示すステップパラメータである。

$\beta_{rt}$  は、時間変化における評価者の厳しさの変化を表すため、 $\beta_{r(t-1)}$  に基づいて  $\beta_{rt}$  を決定する。そのため上記のように仮定する。ただし、 $N(\mu, \sigma)$  は平均  $\mu$ 、標準偏差  $\sigma$  の正規分布を表し、 $\text{LN}(\mu, \sigma)$  は平均  $\mu$ 、標準偏差  $\sigma$  の対数正規分布を表す。既存モデルとの関係から、 $\sigma$  は基本的に0から1の間、かつできるだけ小さくなるのが理想である。 $\sigma$  を上記のように仮定すると、概ねその条件を満たすため、そのように仮定する。

本研究では提案モデルのパラメータ推定手法としてMCMC法を用いる。パラメータの事前分布は  $\theta_j, d_{rk}, \log \alpha_r, \beta_{rt} \sim N(0.0, 1.0^2)$  とした。ここで、 $N(\mu, \sigma^2)$  は平均  $\mu$ 、標準偏差  $\sigma$  の正規分布を表す。本研究では、MCMCのバーンイン期間は1000とし、1000~2000時点までの1000サンプルを用いる。

## 4 シミュレーション実験

本節では、MCMCアルゴリズムによる提案モデルのパラメータ推定精度をシミュレーション実験により評価する。実験手順は以下の通りである。

1. パラメータの真値を、モデルの分布に従って生成する
2. (1) で生成したパラメータを用いて、データを生成する
3. (2) で生成したデータからパラメータ推定を行う
4. (3) で得られたパラメータ推定値と (1) で生成したパラメータ真値において、平均平方二乗誤差 (RMSE) とバイアスを求める
5. 以上を5回繰り返し実行し、RMSEとバイアスの平均値を求める

上記の実験を、学習者数  $j = 60, 90$ 、評価者数  $r = 10, 15$ 、時間区分  $t = 3, 5, 10$  の場合において行った。カテゴリ数は  $K = 5$  とした。

## 5 実データ実験

本章では、実データの適用を通して、提案モデルの有効性を評価する。

本研究では、34名の被験者に4つのエッセイ課題を与え、そのエッセイを34名の評価者が5段階得点で採点したデータを使用する。このデータに提案モデルを適用させるにあたり、今回は項目  $i$  についての推定を行わないため、データを項目ごとに4つに分類した。また、データを評価者が評価を行なった順番に並べ、時間区分が3, 5, 10におけるデータ  $T$  を自分で作成した。例えば、時間区分が3の時には、一人の評価者が評価を行なったのが早い方から3つにデータを分割し、それぞれ TimeID=1, 2, 3 とした。

本研究では、このデータに対して提案モデルを適用する。

## 6 まとめと今後の計画

本研究では、時間区分ごとの評価者の厳しさにマルコフ性を仮定した新しい項目反応モデルを提案した。また、シミュレーション実験と実データを用いた実験を通して、提案モデルの有効性を示した。今後の課題としては、まずはもっと多くの評価者データを集めて推定を行う必要があるという

こと，それに加え，今後は課題特性も考慮して推定を行いたい．今後はこれらの課題を解決し，本研究と同様にシミュレーション実験と実データ実験を行い，モデルの有用性について検証していきたい．

## 参考文献