

論文の内容の要旨

論文題目 確信度付与法による項目反応モデル

氏名 張 一平

最近とみにコンピュータ・テスト (Computer Based Testing : CBT) に関する研究やその実用化が盛んであるが、このコンピュータ・テストには多肢選択式項目 (Multiple Choice) がよく利用され、その応答方式は正答選択式 (Make One Alternative) が一般的である。本論文では、多肢選択形式 CBT の測定効率と妥当性を CBT の自動採点機能を生かしなから更に向上させるために、紙筆式テストでは実施することが困難な確信度付与法 (Probability Testing) についてその特質を数学的に明らかにし、その優れていることをもって正答選択法の代わりに導入することを提案する。そして、確信度付与法によって得られる確信度データを用いて被験者を順位づけるための連続型項目反応モデルを提示し、さらに、提示されたモデルにおける被験者の特性値と項目パラメタの推定方法、情報関数の導出、項目バンクの作り方、項目バンクから項目を抽出する方法などについて、理論的な整備を行う。

第1章「序論」では、多肢選択形式 CBT に利用可能な諸応答法、コンピュータ・テストの発展状況、および、これまでの学力テストにおいて活用された項目反応モデルについて概観した。また、確信度付与法を多肢選択形式 CBT に適用するための項目反

応モデルについての理論的な整備の必要性を明確にした。

第2章「確信度の観点からみた多肢選択式項目の諸応答法」では、離散データを与える応答法である正答選択法、減点法 (Correction for Guessing)、排除法 (Elimination Scoring)、順序付与法 (Ordering Method)、また、連続データを与える応答法である確信度付与法、の5つの応答法の基本型について、確信度の観点から検討を行った。まず、混乱区間 (正答肢に対する確信度と得点の順序関係が逆転する危険性のある正答確信度区間) という全く新たな概念を数学的に定義することによって、5種類の応答法を比較検討した。

各応答法における混乱区間の大きさは、 r を選択肢の数、 P^* を境界確率 (被験者にとってある選択肢が正答であるという主観的確信度が、この値以上でなければ、選択することは利益にならない) として、

$$\text{正答選択法：} \quad S_{OA} = \frac{1}{8} + \frac{[(r-1)/r]^2}{2}$$

$$P^* < 0.5 \text{ の場合の減点法：} \quad {}_1S_{CG} = \frac{3}{8} + \left(P^* - \frac{1}{2}\right)^2$$

$$P^* > 0.5 \text{ の場合の減点法：} \quad {}_2S_{CG} = \frac{1 - (P^*)^2}{2}$$

$$\text{排除法：} \quad S_{ER} = \frac{1}{4} + \left(\frac{r-2}{2r}\right)^2$$

$$\text{順序付与法：} \quad S_{OD} = \frac{23}{192} + \frac{3}{4} \times \left(\frac{5r-12}{12r}\right)^2$$

となる。確信度付与法の混乱区間は r の値に関わらず常に 0 であることも証明され、確信度付与法が優れた方法であることが数学的に明確にされた。

第3章「確信度付与法のための項目反応モデルの提案」では、確信度付与法によって得られる確信度データを利用するために、新たな連続型項目反応モデルを提案した。提案されたモデルは、

$$Z = \Phi^{-1}[I_s(p, q)]$$

$$f(Z|\theta) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{(Z-\rho\theta)^2}{2(1-\rho^2)}\right]$$

と書くことができる。ただし、 Z は項目の確信度得点 S から変換した標準正規得点であり、項目の解答にかかわる能力の高さを表すものである。 Φ^{-1} は累積標準正規分布関

数の逆関数である。 $I_S(p, q)$ は得点 S 以下の被験者の割合を示す不完全ベータ関数比であり、 p と q はベータ分布のパラメタである。 θ はテストの全ての項目に共通する潜在特性である能力変量である。 ρ は Z と θ の相関である。

これはベータ分布で表される確信度分布上のパーセンタイルを正規分布上のパーセンタイルに単調変換することによって、確信度データと共通潜在特性 θ とを一意的に対応させるモデルを意味している。なお、このモデルは共通能力として1因子を仮定したモデルであるということもできる。

本論文で提案した連続型項目反応モデルは、各項目によって測られる能力について正規分布を仮定しており、これまでの項目反応モデルと比べてモデル自体が簡潔で、わかり易いという特徴がある。特性値を推定することも容易で、推定の精度も得られた最尤推定値の密度分布から直接求めることができる。また、ベイズ推定法を使用する場合も、各密度分布がすべて正規分布であるため、複雑な近似計算を行う必要がなく、容易に特性値 θ の事後平均と推定の精度を求めることができる。さらに、これまでの項目反応モデルには必ず存在していた項目困難度パラメタがこの反応モデルでは存在しない。代わりに使用した項目得点の分布のパラメタも、最初の得点変換のときにのみ利用して、特性値の推定式には含まれない。よって、特性値 θ のレベルに関係なく、すべての被験者に対して同じ精度で測定することができる。項目情報量が特性値に依存せず常に項目の識別力の平方に等しいことは、連続型反応モデルの最大の利点である。理論上、提案した連続型項目反応モデルを確信度データに用いる場合の項目情報量は、2値データに2PLモデル(Two-Parameter Logistic Model)を用いる場合の項目情報関数の最大値($0.7225a_j^2$)より3割以上も大きいことが判明した。また、2値データに2PLモデルを用いる場合、コンピュータ・適応型テストを利用しても、被験者に実施したすべての項目に対して最大情報を満足させることは極めて困難である。さらに、他のモデルを用いる場合には、2PLモデルよりも推定精度はさらに低くなる。これらのことを考えると、確信度データに対して提案された連続型項目反応モデルを用いることの情報量に与える効果は大きい。

第4章「連続型項目反応モデルについての検討」では、確信度データに対して提案された連続型項目反応モデルを用いる場合の効果を確認するために、シミュレーション実験を行った。各シミュレーション実験では、すべての項目は各々4つの選択肢を

もつものと仮定された。4.1 節では、2 値データに現在最も多く適用される 2PL モデルを用いる場合の最大推定精度と比べ、確信度データに連続型項目反応モデルを用いる場合の推定精度が、どの程度向上するかについての比較検討を行った。同じ特性値 $\theta = 0$ を持つ 10000 人の人工データセットを用いて、連続型項目反応モデルと 2PL モデルのそれぞれのもとでの最尤推定値とベイズ推定値を求めた。その結果、2 値データに 2PL モデルを用いて推定値を求める場合に最大の項目情報量が得られるような条件下でも、確信度データを利用する場合の方が明らかに高い推定精度が得られることが判明した。使用した項目数が多いほど、2 つの推定精度の差は小さくなるが、推定値の RMSE (Root Mean Square Error) 値の比率は、0.85 を超えることがない。4.2 節では、コンピュータ・適応型テスト形式を利用する場合において、2 値データに 2PL モデルを用いる場合と、確信度データに提案された連続型項目反応モデルを用いる場合のそれぞれの推定精度について比較検討を行った。その結果、2 値データと対応する確信度データを項目の確信度得点の分布の 15 パーセンタイルと 85 パーセンタイルに固定した不利な条件下でも、確信度データからの推定情報量は 2 値データからの推定情報量より大きいという結果が得られた。また、情報量の理論値に近いように確信度データを調整したデータセットの結果から見ると、項目数が少ない場合を除けば、確信度データからの推定情報量は 2 値データからの情報量の約 1.5 倍となることが示された。また、2 値データによる推定の場合、たとえ適応型テストを使用しても、被験者に提示されるすべての項目に最大情報をもたせるのは明らかに困難であることも考えあわせるならば、現実にはシミュレーションの結果より確信度データの利用の効果はさらに大きいものであると考えられる。

第 5 章「確信度付与法と連続型項目反応モデルの応用について」では、確信度付与法を使用する多肢選択形式 CBT の授業や個人指導などへの応用、項目内容と項目バンクの作り方、項目バンクから項目を抽出する方法など、正答選択法を使用する場合と異なる点について理論的考察を行った。

第 6 章「まとめ」では、論文の総括的結論について述べた。本論文では、多肢選択形式 CBT にとって最適な応答法であると考えられる確信度付与法による確信度データを利用するための連続型項目反応モデルを提案したが、今後、理論的な整備に加え、実際に適用する上でのインターフェイスを工夫する必要があることが論じられた。