

## 科学研究費助成事業 研究成果報告書

平成 28 年 6 月 16 日現在

機関番号：32689

研究種目：若手研究(B)

研究期間：2014～2015

課題番号：26780361

研究課題名(和文) Best-Worst尺度法を利用した展開型項目反応モデルによる態度測定法の研究

研究課題名(英文) Proposal of Unfolding item response model using best-worst scaling

## 研究代表者

池原 一哉 (Ikehara, Kazuya)

早稲田大学・グローバルエデュケーションセンター・助手

研究者番号：60706018

交付決定額(研究期間全体)：(直接経費) 1,100,000円

研究成果の概要(和文)：本研究では、複数の選択肢の中から自分の考えや好みに最も合うものと最も合わないものを両方とも選ぶBest-Worst尺度法を利用し、かつ、展開型のメカニズムを考慮した項目反応モデル(BWUモデル)の提案を行った。また、最も合うものを選択するBest尺度法を利用した展開型モデル(BUモデル)も定式化した。シミュレーション研究より、両モデルにおいて適切に母数推定できることが確認され、BWUモデルの方が推定精度は高いことが示された。さらに、CATを実施するために必要な知見を得るための項目提示に関するシミュレーションを行った。

研究成果の概要(英文)：We proposed an unfolding item response model using best-worst scaling (BWU model), in which a person chooses the best and worst stimulus among repeatedly presented subsets of stimuli. We also formulated an unfolding model using best scaling (BU model), and compared the accuracy of estimates between the BU and BWU models. A simulation experiment showed that the BWU model performed much better than the BU model in terms of bias and root mean square errors of estimates. With reference to Usami (2011), the proposed models were applied to actual data to measure attitudes toward tardiness. Results indicated high similarity between stimuli estimates generated with the proposed models and those of Usami (2011). In addition, we developed the algorithm for computerized adaptive test based on BWU model.

研究分野：教育心理学

キーワード：項目反応理論 展開型 態度測定 官能検査 MCMC

## 1. 研究開始当初の背景

項目反応理論は、テストデータや質問紙データを分析して、項目の特徴を捉え、被験者の能力や性格などの潜在特性を測定するための数理モデルである。学力テストでは、特性値の高い回答者ほど難しい項目に正答する確率が高く、また性格検査では、潜在特性の値が高いほど上位カテゴリを選択する確率が高くなる。このように、潜在特性 $\theta$ の値が高くなるにつれて、反応(正答)確率 $P(\theta)$ が高くなるようなモデルは、累積型(cumulative)と呼ばれる。例えば1母数ロジスティックモデルでは、項目の困難度 $b$ に対して潜在特性 $\theta$ が正に大きくなるほど、項目に正答する確率が高くなるようにモデル化される( $P(\theta) = f(\theta - b)$ )。

一方、意見・政策に対する態度の測定や味覚の官能検査では、自分の考えにより近い意見・政策に賛成と反応する確率が高く、また自分の好みにより合った製品においしいと反応する確率が高くなる。例として、甘さの異なるケーキを試食して美味しいか否かを回答する場面を想定する。より甘いケーキに対して、甘党の回答者は美味しいと評価するが、甘いのが苦手な回答者は、美味しくないと評価するだろう。逆に、甘さ控えめなケーキに対して、甘党の回答者は美味しくないと答え、甘いのが苦手な回答者は美味しいと答えるだろう。このように、 $\theta$ の高さではなく、潜在特性 $\theta$ と位置母数 $b$ との距離によって反応確率 $P(\theta)$ が定式化されるモデルは、展開型(unfolding)と呼ばれる。展開型モデルでは、潜在特性 $\theta$ が位置母数 $b$ から離れるほど反応確率が低くなるようにモデル化される( $P(\theta) = f((\theta - b)^2)$ )。

これまで、態度測定や官能検査場面を想定した様々な展開型項目反応モデルが提案されてきた。2値データに適用されるモデル(Andrich, 1988)や順序カテゴリカルデータに適用されるモデル(Roberts, Donoghue, & Laughlin, 2000)の他にも、項目や意見に関する一対比較データのためのモデル(Andrich, 1989)も提案されている。また、多次元展開法の観点から多肢選択データに対するモデル化も行われている(Takane, 1996)。しかし、調査対象となる複数の意見や製品の集合から、一部を取り出して部分集合を作成し、その中に含まれる意見や製品から最も自分に合うものを選択させるモデルは提案されていない。

複数の選択肢の中から最も好ましいものを選ぶ方法はBest尺度法(Best Scaling)と呼ばれる。この手法は多項ロジットモデルとして表現でき、マーケティング分野において、ブランドや商品の選択をモデル化する際に利用されている。談・高野・岸野(2000)では、

不完備デザインを利用した4肢択一モデルと一対比較モデルの比較を行い、Best尺度法の有効性を検討している。

一方、選択肢の中から、自分の考えや好みに最も合うもの(Best)と最も合わないもの(Worst)を両方とも選ぶ方法はBest-Worst尺度法(Best-Worst Scaling)と呼ばれる。Best-Worst尺度法はThurstoneの一対比較モデルの多肢選択への拡張として、Louviere & Woodworth(1990)により提案され、さまざまな分野で利用されている。この手法の利点として、各設問が評価者の対象に対する順位付けに関する情報をより多く含んでいることと、複数の対象の順序付けよりも比較的判断が容易であることが挙げられる。

心理学をはじめとした行動科学の分野では、展開型の回答メカニズムを仮定した潜在特性の測定がよく行われる。例えば、死刑制度や原子力政策、人工妊娠中絶に関する分析では展開型のモデルが利用される。自分の意見と類似した意見・政策には肯定・賛成と回答し、自分の意見よりも極端に革新的あるいは保守的な意見・政策に対しては否定・反対と回答する。このような場合には、潜在特性 $\theta$ の高低は、「肯定—否定」や「革新—保守」のような態度を表すことになる。

従来の態度測定の調査では、評定尺度法が頻繁に利用されており、Best-Worst尺度法を用いた測定はあまり行われていない。項目反応理論においても、複数の選択肢の中から1つを選ぶ名義反応モデルが提案されているが、この手法は項目ごとに選択肢に関する考察を行うためのモデルである。Best-Worst尺度法は、各設問の情報量が多く、判断が容易なため、推定の精度を高めることが期待でき、また、評価者の回答の負担を軽減することが可能である。

## 2. 研究の目的

本研究では、Best尺度法およびBest-Worst尺度法を利用して、効率的にかつ精度よく潜在特性 $\theta$ と項目(意見)の位置母数 $b$ を推定する展開型項目反応モデルの提案を行うこと、そして、提案モデルの応用に関する研究を行うことを目的とする。提案モデルに関して情報量関数を導出し、推定精度の観点から手法の有効性を確認する。また、各設問においてどのような組み合わせで項目(意見)を提示するかについての研究も行う。これにより、回答者の回答を逐一分析して適切な設問を提示する適応型テスト(CAT, computerized adaptive testing)を実施するために必要な知見を得る。

## 3. 研究の方法

はじめに、Best尺度法による展開型項目反

応モデル(BUモデル)と Best-Worst 尺度法による展開型項目反応モデル(BWUモデル)を、多項ロジットモデルによって定式化する。次に、シミュレーション研究による母数推定の妥当性検証を行う。続いて、「遅刻」に対する態度測定の調査を実施し、提案モデルを適応する。調査は関東圏内の大学に通う学生を対象に行い、質問紙を利用してデータを収集する。また、どのような組み合わせで項目を提示することでより精度よく潜在特性を推定できるかについて検討を行い、今後のCATシステムの開発に生かす知見を得る。なお、母数推定には、マルコフ連鎖モンテカルロ法(MCMC法)によるベイズ推定を利用し、サンプリング手法には、ハミルトニアンモンテカルロ法(HMC法)を用いる。

シミュレーション研究では、検証項目として、回答者数( $I$ )、対象の数( $H$ )、設問数( $J$ )の3要因を取り上げる。回答者数は、100人、200人、400人の3通り、対象の数は10個と20個の2通り、設問数は10問と20問の2通りとし、すべての組合せ12通りに関して推定精度の評価を行う。各条件においてシミュレーションデータを50個発生させ、BUモデルとBWUモデルを適用し、推定精度を検証するために、位置母数 $b$ に関してはバイアスとRMSEを、潜在特性値 $\theta$ に関しては相関の平均を算出する。

「遅刻」に対する態度測定の調査では、Usami(2011)を参考に、20個の項目(意見)に関してBest尺度法およびBest-Worst尺度法によってデータを収集し、提案モデルを適用する。提案手法により得られる位置母数の推定値とUsami(2011)の推定値との比較を行い、提案手法の有用性について検証する。

項目の組み合わせの影響度合いの検討においては、3つの提示ルールを用いた場合の潜在特性値の推定の精度を比較する。

今後の展開として、展開型モデル(理想点モデル)を角度を用いて分析することを考える。そのために、角度データをMCMCで分析する際の収束判定に関する検討を行う。

#### 4. 研究成果

##### (1) シミュレーション研究

位置母数と潜在特性値のシミュレーション結果を表1に示す。表1より、BUモデルおよびBWUモデルに共通して、回答者数が多くなるほど、バイアスの値やRMSEの値が小さくなる傾向にあることが分かる。また、対象の総数が同じであれば、設問数が多いほどRMSEの値が小さく、精度良く推定できることが示された。なお、潜在特性値に関する相関は、どの条件においても0.95以上であり、提案モデルにより真値を復元できることが分かった。

BUモデルとBWUモデルを比較すると、BWUモデルの方がバイアスやRMSEの値が小さく、相関の値が大きいことが示された。さらに、 $H = 10$ の場合に注目し、設問数20

のBUモデルと設問数10のBWUモデルを比べると、位置母数に関するRMSEや潜在特性値に関する相関が同程度であることが確認できる。つまり、設問数を減らしてもBWUモデルを適用することで同程度の精度で母数推定を行えることが示唆される。なお、同様な傾向が $H = 20$ の場合においても見られた。

表1 各母数におけるシミュレーション結果

$I$		$b_{bias}$			
		H10J10	H10J20	H20J10	H20J20
100	BU	0.0147	0.0125	0.0214	0.0154
	BWU	0.0142	0.0119	0.0158	0.0117
200	BU	0.0078	0.0080	0.0150	0.0073
	BWU	0.0051	0.0053	0.0103	0.0072
400	BU	0.0089	0.0085	0.0095	0.0052
	BWU	0.0061	0.0073	0.0064	0.0043
		$b_{RMSE}$			
100	BU	0.1334	0.1086	0.1584	0.1194
	BWU	0.1097	0.0972	0.1221	0.1050
200	BU	0.0882	0.0771	0.1115	0.0838
	BWU	0.0783	0.0679	0.0840	0.0687
400	BU	0.0688	0.0583	0.0832	0.0638
	BWU	0.0566	0.0532	0.0637	0.0542
		$\theta_{cor}$			
100	BU	0.9616	0.9781	0.9551	0.9771
	BWU	0.9763	0.9871	0.9721	0.9859
200	BU	0.9599	0.9783	0.9559	0.9769
	BWU	0.9757	0.9873	0.9735	0.9858
400	BU	0.9611	0.9785	0.9563	0.9766
	BWU	0.9757	0.9873	0.9735	0.9860

##### (2) 実データの分析

大学生・大学院生172名に調査を実施し、BUモデルとBWUモデルを適用した。Usami(2011)の遅刻に関する20個の意見を、各設問において4つずつ提示し、最も自分の考えに合うものと最も合わないものを選択してもらった。設問パタンの作成には統計解析環境Rの関数find.BIBを利用した。

BUモデルおよびBWUモデルによって位置母数を推定し、Usami(2011)の推定値とのピアソンの積率相関係数を計算したところ、BUモデルでは0.965、BWUモデルでは0.970と非常に高い値となった(ケンドールの順位相関係数はそれぞれ0.842と0.853であった)。

また、BUモデルとBWUモデルの推定結果を比べると、BUモデルの方が位置母数の分散が大きく、広範囲に散らばっていることが分かった。なお、事後標準偏差の平均はBWUモデルの方が小さく、Best-Worst尺度法を利用することで推定精度が向上することが示唆される。

続いて、潜在特性値のEAP推定値のヒストグラムを図1(BUモデル)と図2(BWUモデル)に示す。ヒストグラムには濃い黒い線で位置母数も同時に示した。図1と2より、

BUモデルの潜在特性値の方がBWUモデルよりも散らばりが大きく、BWUモデルでは0付近の度数が多くなっていることが分かる。潜在特性値は、両極端の意見の位置母数の内側に分布しており、-1から1の間にほとんどの回答者が含まれている。

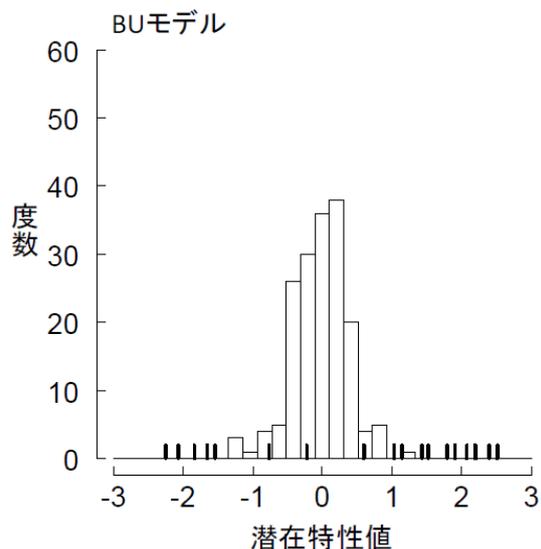


図1 BUモデルのヒストグラム

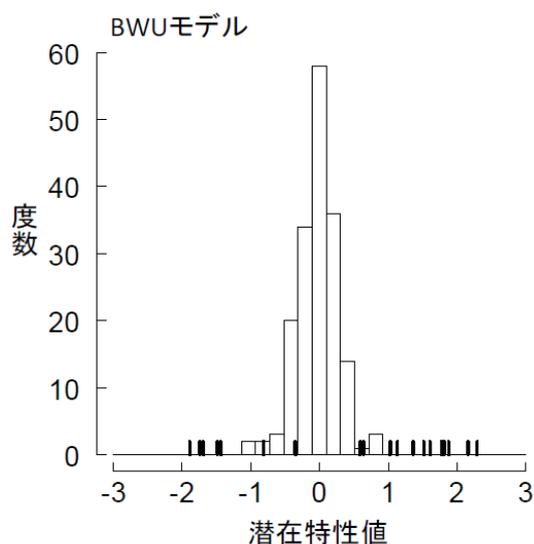


図2 BWUモデルのヒストグラム

最後に、BWUモデルの推定値を利用して情報量を計算し、潜在特性のレベルごとの $SD(\theta)$ をプロットした図を図3に示す。図3には、両モデルにおいて、20個すべての設問を利用した場合の $SD(\theta)$ の曲線(BU20とBWU20)と前半10個の設問を利用した場合の $SD(\theta)$ の曲線(BU10とBWU10)をプロットした。

図3より、設問数によらず、BWUモデルの方が、BUモデルよりも潜在特性のレベルごとの $SD(\theta)$ が小さく、精度が良いことが分かる。また、設問数が増えることで $SD(\theta)$ が小さくなり、推定精度が向上することも確認できる。さらに、-0.3から0.4の間では、

20個の設問を用いてBUモデルを適用するよりも、10個の設問を用いてBWUモデルを適用したほうが精度がよく、0.4から1.5の範囲においても同程度の誤差で潜在特性値を推定できることが示唆された。

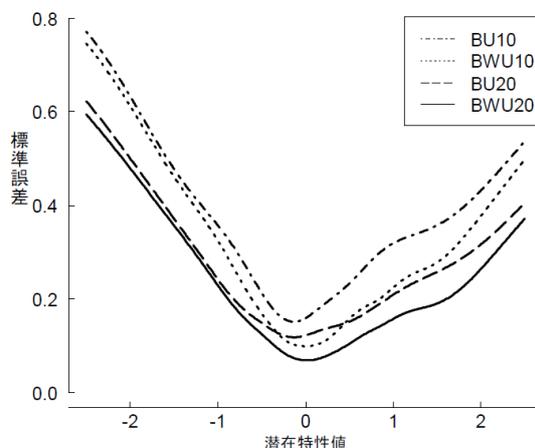


図3 潜在特性の標準偏差 $SD(\theta)$

### (3) 項目提示ルール of 検討

項目の提示ルールを検討するために、①ランダムに提示する項目を選ぶ方法、②潜在特性値と位置母数の距離を利用して確率を規定し、その確率によって項目を選ぶ方法、③位置母数の値の大きさで4グループ(四分位点分割)に項目を分割し、各グループから1つずつ選ぶ方法の3つに関して、シミュレーションデータにより検討を行った。

位置母数に関しては、標準正規分布から発生させて、40項目のプールを作成した。また、位置母数の値の四分位点を計算し、グループ1(0%から25%点まで)、グループ2(25%から50%点まで)、グループ3(50%から75%点まで)、グループ4(75%点から100%点)の4つのグループを作成した。最初に提示する共通の設問の数を5とし、4つのグループからそれぞれ1つずつ項目を選ぶように設問を構成した。潜在特性値に関しては、真値を-1、0、1の3つに設定し、各点で100回分のデータを発生させた。40回設問が提示された段階で終了し、設問数 $J = 10, 20, 30, 40$ のときの結果を比較する。なお、推定精度の評価には、MSEを利用した。

シミュレーションの結果を表2に示す。表2より、項目数が増えるにつれて精度が向上していることが分かる。3つの方法を比較すると、潜在特性の値が0.0のときは四分位点分割による提示方法がより精度が高いことが示された。一方、潜在特性の値が-1や1のときは、それほど大きな違いは見られなかった。提示方法に関しては、潜在特性の値に応じて選択方法を変えることも一つの案として考えられる。

表 2 提示項目のシミュレーション結果

ランダム提示			
	-1.0	0.0	1.0
$n = 10$	0.0275	0.0132	0.0210
$n = 20$	0.0190	0.0087	0.0088
$n = 30$	0.0087	0.0042	0.0044
$n = 40$	0.0049	0.0051	0.0030
距離による確率による提示			
	-1.0	0.0	1.0
$n = 10$	0.0158	0.0180	0.0178
$n = 20$	0.0093	0.0085	0.0087
$n = 30$	0.0066	0.0065	0.0041
$n = 40$	0.0052	0.0031	0.0029
四分位点分割による提示			
	-1.0	0.0	1.0
$n = 10$	0.0178	0.0144	0.0222
$n = 20$	0.0073	0.0014	0.0121
$n = 30$	0.0030	0.0000	0.0099
$n = 40$	0.0017	0.0000	0.0095

(4) 角度データの収束判定

今後の展開として、展開型モデル（理想点モデル）を角度を用いて表現することを想定し、角度データを MCMC で分析する際の収束判定の研究を行った。

角度データは、定義域が  $0 \sim 2\pi$ 、もしくは  $-\pi \sim \pi$  となるため、MCMC による収束判定を行う際には、注意が必要となる。検証項目として、サンプルサイズ ( $N$ )、位置母数 ( $\mu$ )、集中度母数 ( $\kappa$ ) の 3 つを取り上げた。サンプルサイズは  $N = 25, 50, 75, 100$  の 4 通り、位置母数は、 $\mu = \pi/4, \pi/2, 3\pi/4$  の 3 通り、集中度母数は  $\kappa = 0.5, 0.7, 1.0, 3.0, 1.0$  の 5 通りとし、すべての組み合わせ 60 通りに関して収束判定の振る舞いを検討した。なお、収束判定には、Rhat、Geweke 指標を利用した。

シミュレーションの結果、位置母数の収束判定には、Rhat や Geweke 指標をそのまま利用するのではなく、 $\cos$  変換や  $\sin$  変換を施した値に対して Rhat や Geweke 指標を用いることがより適切であることが示唆された。

5. 主な発表論文等

〔雑誌論文〕 (計 1 件)

- ① 池原 一哉、Best-Worst 尺度法を利用した展開型項目反応モデル、心理学研究、査読有、第 85 巻、第 6 号、2015、560-570

〔学会発表〕 (計 2 件)

- ① Ikehara, K. Unfolding item response model using best-worst scaling: measurement of attitude toward tardiness. International Meeting of Psychometric Society 2014、2014 年 7 月、Madison (United States of America)

- ② Ikehara, K. & Akiyama, T. Comparison of Behavior of Convergence Diagnostics for Markov Chain Monte Carlo Methods When Analyzing the Circular Data. East Asia Regional Biometric Conference 2015、2015 年 12 月、Kyushu University (Fukuoka)

6. 研究組織

(1) 研究代表者

池原 一哉 (IKEHARA, Kazuya)

早稲田大学・グローバルエデュケーションセンター・助手

研究者番号：60706018

(2) 研究分担者

なし

(3) 連携研究者

なし