

多肢強制選択型心理測定の回答負荷を 軽減するための項目提示法

○ 分寺杏介¹ 杉山剛²

¹ 神戸大学経営学研究科 ² 株式会社リクルートマネジメントソリューションズ

1 問題と目的

リッカート尺度におけるバイアス等の問題への対抗策として、1つの項目内で複数の文を選択肢として同時に提示し、自分を最もよく表している文を選ばせる多肢強制選択型 (Multiple alternative forced choice; 以下 MAFC) の心理測定が近年注目を集めている。しかし、MAFCでは回答者に複数の文を解釈し、比較を行うことを要求するため、回答に必要な認知負荷が非常に高いという問題がある (Bürkner et al., 2019)。この負荷が正確な回答を妨げている可能性があるため、本研究では MAFC における回答負荷を軽減するための新たな項目提示法を提案し、実際にデータを収集してその効果を検証する。

2 項目提示法

以下では、項目に含まれる文 (選択肢) の組み合わせを「ブロック」と呼ぶ。また本研究では、各ブロックに対して「自分を最もよく表している文を1つのみ選ぶ」回答形式を取り扱う。MAFCでは通常、1つのブロックへの回答が完了すると、次のブロックはすべて新規の文で構成される。これが1つの回答を決定するために必要な認知負荷を高める要因であると考えられる。そこで、次のブロックを構成する際に、前のブロックで選択されなかった文はそのまま使用する手法を提案する。この場合、前のブロックで選択された文のみが新規の文に置き換わるため、新規に提示される文の数はブロックごとに1つのみとなり、認知負荷が軽減されることが期待される。表1に、仮想的な回答パターンおよび提案手法に基づいて各ブロックに提示される選択肢を示した。例えばブロック1で文Bを選択した場合、残る文A, C, Dはブロック2にも続けて表示され、選択された文Bに変わって新たに文Eが提示される。なお新規に提示される文は、通常のMAFC形式の尺度に合わせて、残された3つの文とは別の因子から選択する。

3 データ拡張

提案した項目提示法では、1つの回答ごとに提示される新規の文は1つのみのため、従来の方法 (一般的なMAFC) と比べて、回答数が同じ場合には提示される文の総数は減少する。このことは、項目パラメータの推定には良い影響を与える可能性がある一方で、回答者の因子得点の推定には悪

表1 仮想的な回答パターン（下線は選択された文を表す）

ブロック	選択肢 1	選択肢 2	選択肢 3	選択肢 4
1	文 A	<u>文 B</u>	文 C	文 D
2	文 A	文 E	文 C	<u>文 D</u>
3	文 A	文 E	文 C	<u>文 F</u>
4	文 A	文 E	<u>文 C</u>	文 G

影響をもたらす可能性が考えられる。この問題に対応するため、項目提示法に合わせた回答データの拡張法を提案する。提案する手法では同じ文が複数回提示されているため、観測された回答パターンに基づき、各文の選好の大小関係に推移律が満たされていると仮定して回答パターンを推測することができる。以下では、表1をもとに、具体的なデータ拡張法を説明する。MAFC形式によって得られた回答データの分析では、得られた回答を仮想的な二値データに変換し、Thurstonian IRTモデル（以下 TIRT; Brown & Maydeu-Olivares, 2011）を使用することが一般的である。

1. ブロック 1 の回答から、各文の選好の大小関係について「文 B > 文 A, C, D」と推測される。
2. 同様にブロック 2 の回答からは「文 D > 文 A, C, E」と推測される。このとき、初回の回答と合わせると「文 B > (文 D) > 文 E」と推測される。この場合、実際に直接比較は行われていないが「文 B > 文 E」という回答が得られたものとしてデータを拡張する。
3. ブロック 3 では、新規に提示された文 F が選択されている。この場合、文 F と文 B, D の間の大小関係を推測することはできないため、データの拡張は行わない。

4 実データ解析

提案手法およびデータ拡張の効果を検証するため、Web 調査にて実データを収集し、解析を行った。Web アンケートモニター会社を介して調査への協力を依頼し、以下の 3 つのフェイズに分けてそれぞれ異なる方法でデータ収集を行った。協力者はフェイズ 1 から順に調査に参加した。なお所要時間の関係上、フェイズ 3 はフェイズ 1, 2 とは別日に実施した。

フェイズ 1 画面上に 4 つの文が提示され、協力者はその中から「自分を最も表しているもの」を選択した。回答決定後は、提案手法に基づいて次のブロックが提示され、ブロックが構成できなくなるまでこれを繰り返した。なお調査で使用した文は、リクルートマネジメントソリューションズ社が保有する性格検査の質問文をもとに、シミュレーション用に作成した 272 文であった。

フェイズ 2 画面上に 4 つの文が提示され、協力者はそれらを「自分をよく表すと思う」順に選択した。提示する文およびブロックは、フェイズ 1 で使用したものの中から 32 文を選択し、筆者らが事前に構成した計 8 ブロックを使用した。

フェイズ 3 フェイズ 1 で使用した全ての文を 1 つずつ提示し、協力者は「あてはまる」から「あてはまらない」の 4 件法で回答を行った。

最終的に、フェイズ 1, 2 の両方に回答した協力者は 1260 名、またフェイズ 3 まで回答した協力者は 786 名であった。このうち、(a) ストレートラインの最大値が 30 以上、(b) フェイズ 1 におけ

る平均回答時間が 400ms 未満, (c) フェイズ 1 のうち 3 ブロック以上で回答時間が 400ms 未満, (d) フェイズ 3 まで回答した協力者についてはフェイズ 3 での平均回答時間が 2000ms 未満, のいずれかを満たす協力者のデータを除外し, 最終的にフェイズ 1, 2 については 816 名, フェイズ 3 については 577 名のデータを使用して分析を行った。

■**モデルおよび設定** 項目パラメータおよび特性値の推定は, stan を用いて行った。尤度関数は

$$P(x_{jk} = 1 | \boldsymbol{\eta}) = \Phi \left[\frac{(\mu_j + \beta_j \eta_a + \nu_j) - (\mu_k + \beta_k \eta_b + \nu_k)}{\sqrt{\Psi_j^2 + \Psi_k^2}} \right], \quad (1)$$

また事前分布は

$$\begin{aligned} |\beta_j| &\sim \text{normal}(1, 3), & \mu_j &\sim \text{normal}(0, 3), & \Psi_j &\sim \text{normal}(1, 0.3), \\ \nu_j &\sim \text{normal}(0, \Psi_j), & \boldsymbol{\eta} &\sim \text{multi_normal}(0, \boldsymbol{\Sigma}), & \boldsymbol{\Sigma} &\sim \text{lkj_corr}(1), \end{aligned} \quad (2)$$

と設定した。なお, x_{jk} は文 j および文 k の比較において文 j が選択された場合に 1, 文 k が選択された場合に 0 となる二値変数, (β, μ, Ψ^2) はそれぞれ項目の因子負荷, 切片, 誤差分散を表している。 ν は, stan での推定の際に各データポイントの尤度を条件付き独立にするために追加すべき誤差項 (Bürkner et al., 2019) である。また, 回答者の性格特性は η で表されている。

結果

■**因子構造** TIRT モデルによるパラメータ推定は確認的因子分析の枠組みで行われるため, 各文がどの因子を反映しているかは所与とされる。そこで本研究では, フェイズ 3 の回答データを用いて探索的因子分析を行った。その結果, 9 因子構造が妥当であると判断した。以後は, この探索的因子分析で得られた因子負荷行列を, 因子負荷の絶対値が最大の 1 因子のみを残して完全単純構造に変換したうえで分析に使用する。

■**データ拡張の精度** 本研究が提案するデータ拡張法の有効性は, 実際にその文のペアが同時に提示された場合と同じ回答をするか (3 節の例では, 実際に文 B と文 E の比較を行った場合には文 B のほうが選択されるか) の精度によって決定すると考えられる。そこで, フェイズ 1 の回答データから作成した拡張データと, フェイズ 2 の回答データを比較し, 同一の文のペアが存在するものについてその一致率を確認した。その結果, 2 つのパターンで回答が一致したものは約 77.89% (597 件中 465 件) であった。一方, フェイズ 1 の回答データの中に存在する比較対とフェイズ 2 の回答データを比較した場合, すなわち同一の協力者が実際に同じ比較対に 2 度回答した場合の回答の一致率はおよそ 80.80% (224 件中 181 件) であった。2 つの一致率に有意差が見られなかった ($\chi^2(1) = 0.660, p = .4165$) ことから, 提案手法によるデータ拡張によって, 実際に回答した場合と同程度の精度の回答データが作成できたといえる。

■**回答時間** 提案した項目提示法によって回答負荷が軽減されているかを評価するため, フェイズ 1 とフェイズ 2 で得られた回答データについて, 回答時間の比較を行った。フェイズ 1 では, 各ブロックにおいて 1 位の文を回答するまでの時間が記録された。一方フェイズ 2 では, 4 つの文について完全なランキングを決定するため, 3 位の文までを順に選択する合計時間が記録された。2 つのデータにおける回答時間を厳密に比較するためには, フェイズ 2 で記録された回答時間のうち

「1位の文を回答するまでの時間」を評価すべきだが、システム上不可能であったため、便宜的にフェイズ2で記録された回答時間を3で割った値を使用した。ブロックごとに記録された回答時間(フェイズ2については3で割った値)を対数変換した値の個人内平均値を算出し、その平均値を計算した後で元のスケールに戻したところ、フェイズ1での平均回答時間は約2867ms、フェイズ2での平均回答時間は3172msであった。2つのフェイズにおける差はさほど大きくないようも見えるが、実際には、1位の回答を決定するまでの時間が最も長いと考えられるため、フェイズ2における回答時間は今回使用した値よりも長い可能性が高い。以上の結果から、提案手法によって1ブロックの回答に必要な認知負荷は軽減できたと考えられる。

■因子得点 データ拡張によって因子得点 η の推定に情報を補完できるかを確認するため、フェイズ1の回答データについて、データを拡張せず観測された結果のみを用いて推定した場合と、データ拡張を含めて推定した場合で得られる結果を比較した。図1の左は、EAPの散布図である。一部において大きく変動したものが見られるが、全体としてはデータ拡張前との相関関係が保たれている($r = .861$)。また図1の右は、事後分布の標準偏差の散布図である。大半の点が $y = x$ の直線よりも下に見られることから、多くの場合においてデータ拡張によって η の事後分布の幅がより狭まっているといえる。

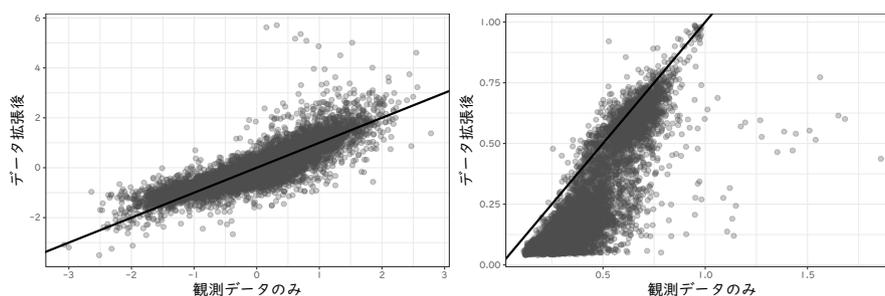


図1 データ拡張前後で推定された η の散布図(左: EAP, 右: 事後SD)

5 まとめ

本研究が提案した項目提示法によって、1ブロックあたりの回答負荷(回答時間)は多少軽減されると考えられる。また、データ拡張を組み合わせることで、回答から得られる情報量の減少にも対応可能であることが示唆された。一方で、データ拡張による η の推定値の変動が、真値と比較してどのような影響であるかについては今後の検討が必要である。また、調査全体での回答負荷については十分な比較ができていないため、こちらについても今後検討を重ねていく予定である。

参考文献

- Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement, 71*(3), 460–502. <https://doi.org/10.1177/0013164410375112>
- Bürkner, P.-C., Schulte, N., & Holling, H. (2019). On the statistical and practical limitations of Thurstonian IRT models. *Educational and Psychological Measurement, 79*(5), 827–854. <https://doi.org/10.1177/0013164419832063>