

調査項目の精選と評価 —項目反応理論による検討—

片山 浩子（岡山理科大学 留学生別科）

要約:

カテゴリカルデータを用いた調査や検査では、しばしば項目数が多いという問題に直面する。そこで、元の項目の様相を保った上でどのように項目を精選することについて検討する。具体的には、カテゴリカルデータを項目反応理論に適用し、推定された潜在特性値の結果を用いて、項目の精選を行う。さらに、実データを用いて項目精選の様相を確認したのち、精選された項目からみた元の質問項目の評価を行うことを検討する。

キーワード: カテゴリカルデータ, 項目反応理論, 潜在特性値, 項目精選, 項目の評価

1. はじめに

カテゴリカルデータを用いた調査や検査は詳細な情報を得るために項目数が多くなってしまふことが課題となっている。調査項目が多いと被験者に負担をかけてしまふことや、特に調査に不慣れな子どもやお年寄りに対する調査について、並川（2012）は、60項目のみでも実施が困難になる場合があるため、調査における負担軽減は研究倫理的な面でも改善が望まれる点であることを指摘している。従来行われている調査票の短縮版の検討には、因子負荷量が高いものの順に項目を選ぶ方法を採用している場合が多くみられるが、この方法では項目の困難度に関する情報が反映されないため、必ずしも十分な検討とは言えないことを懸念している（並川，2012）。そこで、項目の良し悪しや各項目に対する難易度、被験者の潜在特性値が推定できる、項目反応理論（IRT: Item Response Theory）を利用することを考える。この方法は、元来テスト理論で利用される分析方法で、テスト問題の各項目の良し悪しやそれら項目の難易度、被験者の潜在能力を分けて分析することができる特徴をもっている。この理論の特徴を活かし、現在はテストの分析に限らず、心理学など幅広い分野で利用されている。具体的に IRT を利用し、調査票の短縮版の検討を行っている研究には、笹川 他（2004）、脇田（2004）、並川 他（2012）、浦上 他（2016）などの先行研

究がある。笹川 他 (2004) は、心理測定に段階反応モデルを用いて 5 件法の短縮版の検討を行っている。脇田 (2004) は、選択肢の数や、「あてはまる」などの選択肢のカテゴリの表現を変えることは、回答者の反応傾向にも影響を与える可能性があることを指摘したうえで、IRT を用いた短縮版の検討を行っている。並川 他 (2012) は、IRT で推定された情報量を用いて項目精選の検討を行い、浦上 他 (2016) は、IRT を利用することにより、測定次元を変えないで短縮版の検討が行えることを示唆している。並川 (2015) は、調査票の短縮版の作成方法として最も多くの論文で行われているのは、因子分析の結果に関する情報をもとにした項目選択を行う方法であり、IRT を用いた短縮版の検討が行われている論文が少ないことから、IRT を用いた短縮版作成は今後の重要な検討課題であるとしている。

そこで、本研究では、IRT で推定される潜在特性値に注目し、被験者がもっている能力の観点から項目の精選と項目の評価について検討を行う。

2. 項目反応理論を用いた項目精選

2.1 項目反応理論

まず、IRT についてまとめる。IRT は従来克服できなかった集団的依存性と項目の依存性を克服するために新たな評価方法として考案された理論である (加藤 他, 2014)。IRT の特徴は、テストに含まれる項目の難易度と受験者の能力を別々に表現できることにある。通常、あるテストが行われ、集団の平均点が高かったとき、そのテスト項目が簡単だったからなのか、このテストを受けた学生が優秀だったからなのかは、テストの点数を素点や単純集計でみるだけでは判断できない。この問題を解決するために、IRT では、問題の難易度のモデルを受験者 n 人に対して p 個の項目の正答率をもとに推定する。この推定には、受験者の潜在的な能力を θ 、ある項目 j に正答する確率を $P_j(\theta)$ としたとき、一般的に 2 パラメータロジスティックモデルが適用される ($1 \leq j \leq p$)。

$$P_j(\theta) = \frac{1}{1 + \exp[-1.7a_j(\theta - b_j)]}, \quad -\infty < \theta < \infty \quad (2.1)$$

この θ と $P_j(\theta)$ の関係をもって、項目 j を評価しようというものである。この方法を用いることで各被験者の潜在特性値 (θ) を推定し、被験者の普遍的な能力を測ることができる。これらにより、項目の難しさや項目が能力を識別している程度を評価し、また、推定された θ から各項目に対して被験者の能力を評価する。他にも能力が低い受験者でも、偶然正答

する確率である c_j 当て推量を取り入れた3パラメータロジスティックモデルがある。この

$$P_j(\theta) = c_j + (1 - c_j) \frac{1}{1 + \exp(-D_{aj}(\theta - b_j))}, \quad -\infty < \theta < \infty \quad (2.2)$$

(2.2) 式で用いられている D_{aj} の D は尺度因子を示し、一般的には、 $D=1.7$ という値が設定される。 a はすべての項目に共通の値をもつ識別力であり、 j は j 番目の項目である。また、多値の場合に用いる段階反応理論

$$P_{jk}^+(\theta) = \frac{1}{1 + \exp[-a_j(\theta - b_{jk})]}, \quad -\infty < \theta < \infty \quad (2.3)$$

などがある (samejima, 1969)。式 (2.3) の k には例えば、選択肢 1, 2, 3, 4, 5 をあてて、5 種類の確率推定を行うことができる。

IRT を利用するためには条件があり、使用するデータは次元性をもつデータであるのかを確認しなければならない。また、2 値データを利用する場合は、利用するデータが 2 パラメータロジスティックモデルに適合しているのか、または 3 パラメータロジスティックモデルに適合しているのかを確認し、適合が良いほうを利用する。以上のような手順を踏んだのちにカテゴリカルデータの分析に IRT の使用が可能となり、項目の精選に利用できるようになる。

2.2 カテゴリカルデータの IRT への適用

ここでは、能力について、収集したカテゴリカルデータを IRT に適用することを考える。IRT では、識別力 (a_j)、困難度 (b_j)、潜在特性値 (θ)、正答確率 $P_j(\theta)$ が解釈すべきパラメータとして用いられるが、能力について尋ねた調査では、各パラメータを次のように解釈をし直す。識別力は自身の能力を区別する項目と考える。識別力が低ければ自身の能力を区別できない項目であるとし、識別力が高ければ、自身の能力をよく区別できた項目であるとする。困難度は、マイナスであれば、自身の能力に対して肯定的であると捉え、プラスであれば、否定的な考えであることを示す。3 パラメータロジスティックモデルを採用する場合、この時の c_j は、被験者が適当に項目に回答する確率とする。潜在特性値 (θ) は、被験者の普遍的な能力に対する意識とする。今回は、被験者の普遍的な能力 (θ) に対する意識の結果を用いて項目選択を行うため、普遍的な能力 (θ) のみ利用する。

2.3 潜在特性値を使った項目精選

以下のように、IRT を適用して得られた被験者の潜在特性値 (θ) を利用して、項目精選の方法を示す。ここで、IRT で推定された項目 i に対する各個人の潜在特性値の推定結果を (θ_i) とし、これらの結果を利用して、項目選択の方法を考える ($i=1, \dots, n$)。 p 個すべての項目 \mathbf{Y} から推定された被験者 i の潜在特性値を $\theta_{i(p)}$ とし、 q 個の部分項目群 \mathbf{Y}_1 から得られた被験者 i の潜在特特性値を $\theta_{i(q)}$ とする ($2 \leq q \leq p$)。このとき、次の規準 d を考え

$$d_q = \sum_{i=1}^n (\theta_{i(p)} - \theta_{i(q)})^2 \quad (2.4)$$

すなわち、個々の i に対する $\theta_{i(p)}$ と $\theta_{i(q)}$ の差の二乗和である d_q を選択基準として、この d_q を最も小さくする q 個の項目 \mathbf{Y}_1 を選ぶことにより、最適な項目群を見つけようというものである。以下、項目選択の手順を示す。

1. 全ての変数 \mathbf{Y} から推定された被験者 i の潜在特性値 $\theta_{i(p)}$ を求める。
2. 変数の数が q である \mathbf{Y}_1 の全ての組み合わせに対して $\theta_{i(q)}$ を求める。
3. 1 と 2 に対して ${}_p C_q$ 個の d_q を求める。
4. ${}_p C_q$ 個の d_q のうち、最も小さな d_q を提供する \mathbf{Y}_1 を最適な \mathbf{Y}_1 とする。

なお、上記は総当たり法であるため、 p の値によっては、大幅な計算時間がかかるため、実際の選択においては、変数減少法を用いることにした。

3. 数値例

3.1 データ

IRT を利用した項目選択の動作をみるために、職業能力評価および資格の役割に関する調査と在宅介護ヘルパーの仕事と能力に関するアンケート調査データ (東京大学社会科学研究所人材ビジネス研究寄付研究部門) の 2 つのデータを利用する。1 つ目のデータは、自身もつ仕事の能力について尋ねたものである。この調査は、720 名を対象に行い、質問内容は各質問項目の業務に対する能力について長けていると思うかを尋ねた問題になっている。項目数は全部で 70 項目である。質問内容の特徴は、「事務系 (中堅社員)」、「営業系 (中堅社員)」、「技術系 (中堅社員)」、「現業系 (中堅社員)」、「課長」の職業能力について項目が 1~14 項目あるデータとなっている。このデータを IRT に適用するため、該当する「1」、該当しないを「0」とし、2 値データとして扱った。このデータの次元性を確認したのち、データ全体の適合度を見る必要があるため、2 パラメータロジスティックモデルと 3 パラメータロジスティックモデルの情報量を確認した。今回のデータは 3 パラメ

一タロジスティックモデルに適合したため、式 (2.2) を用いて潜在特性値 (θ) を推定する。これをデータ 1 とする。

2 つ目のデータは、ひとりで (同行指導や家族の指示なしに) 食事介助など介護の各分類のそれぞれの仕事がどの程度できるかを調べる調査である。調査人数は 1450 名で、この調査は 18 分類の仕事の能力についてそれぞれ 3 つの仕事内容に関する質問項目が設定され、計 54 項目について回答するようになっている。個別の仕事 (能力) に対して「5.実務経験があり確実にできる」、「4.実務経験がありかなりできる」、「3.実務経験がありだいたいできる」、「2.実務経験があり少しできる」、「1.実務経験なし・ほとんどできない」の 5 段階で測るものである。これをデータ 2 とする。このデータを (2.3) 式に適用し、潜在特性値を推定し、項目精選に利用する。

3.2 項目数について

データ 1 について今回利用したデータは、3 パラメータロジスティックモデルの方が情報量の適合度が良かったので 3 パラメータロジスティックモデルを採用した。データ 1 については以上の手続きをしたのち、2 値データについて式 (2.2) で推定された潜在特性値 (θ) を用いて、式 (2.4) に適用し、項目選択を実行した。項目数の決定については、データ 1 の場合もデータ 2 の場合も同様の方法を用いるため、ここでは、データ 2 の結果のみ示す。

次に項目選択を実行した結果を表 1 に示す。 q の列は選択された数、 $\mathbf{Y}_1 | \mathbf{Y}_2$ の左側が d_q の規準で選ばれた項目番号 \mathbf{Y}_1 、右側が落とされた項目番号 \mathbf{Y}_2 である。 d_q は規準値を示す。ここでは、項目選択の方法として変数減少法を用いた。その d_q の変化を図 1 に示す。

もし、調査者が使用したい項目の具体的な数が決まっていれば、使いたい数の q の行 \mathbf{Y}_1 を用いればよい。具体的な数が決まっていない場合は、 d_q の前後の差 (q 個のときの d_q と d_{q+1} 個のときの d_{q+1} の差 : $d_q - d_{q+1}$) が広がり始めるところで数を決める方法もある。例えば、図 1 の d_q の変化を見ると、 q を 53 個から順に減らしていくと、 d_q の前後の差は、1.08, 0.52, 1.26, 0.20... と基本的には、差が少しずつ大きくなっていく。 q が 27 のとき 1.76, 26 のとき 34.8 と、このあたりから差が開き始める。例えばこのタイミングで、誤差が広がったと判断して $q=27$ のところで選ばれる項目を採用することが考えられる。すると、選ばれた項目は、2, 3, 4, 5, 7, 8, 10, 14, 15, 19, 20, 21, 23, 25, 26, 27, 31, 37, 38, 39, 43, 44, 45, 49, 51, 52, 54 となる。このように d_q の前後の

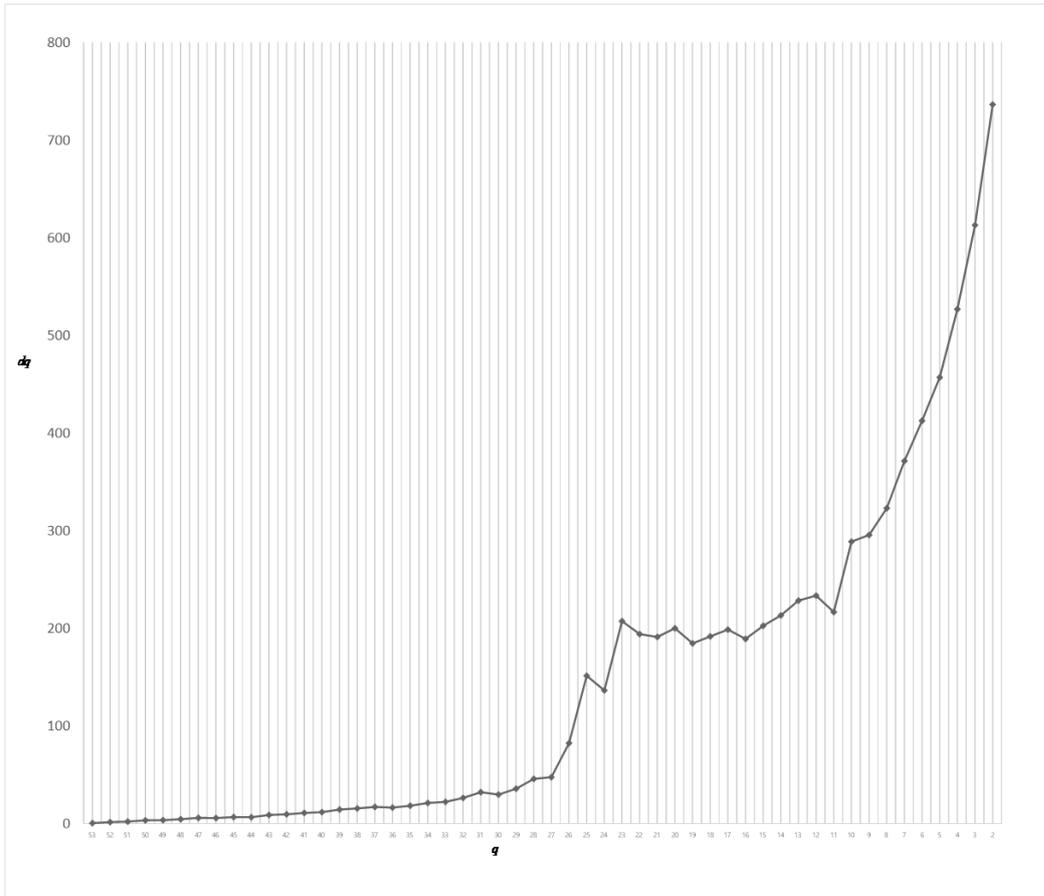


図1 d_q の変化 (データ 2)

3.3 項目の精選と評価

例えば、3.2 節で決めた方法で項目を決定する場合、 $q=27$ の場合をみると、食事介助の項目 2, 3, 排泄介助 4, 5, 更衣介助 7, 8, 入浴介助 10, 清拭 14, 15, 体位変換 19, 20, 21, 移乗介助 23, 外出介助 25, 26, 27, 掃除 31, 健康チェック 37, 38, 39, 説明 43, 44, 45, 情報収集と判断 49, 51, 協働 52, 54 である。このようにして、 d_q の規準から項目の精選が可能となった。

次に項目の評価について考えるため、表 2 に項目が何回選ばれたのか回数を示す。この表は q がいくつになったかのタイミングでどの項目が選ばれたのかを示している。○がついている項目は選ばれた項目を示し、○がついていない項目は選ばれなかった項目を意味している。この表を見ると、項目 11, 項目 32, 項目 42 などは早い段階から落とされており、逆に、項目 7, 項目 15, 項目 19 などの項目は最後まで残っていることがわかる。例えば、この選ばれた項目が重要であるとするならば、これらの質問は全体の潜在特性値 θ の情報をよく保持している質問であり、今後の調査でも必ず用いるべき項目

利用するか 3 パラメータロジスティックモデルを利用するかについての検討が必要であるため、データがもつ情報量で適合度を確認した。今回使用したデータは 3 パラメータロジスティックモデルが適合したので、このモデルを採用した。多値データについては、段階反応理論を適用し、推定された潜在特性値を項目精選に利用した。

その結果、数値例により 2 値データや多値データの場合でも項目選択することが可能であることを示すことができた。これら IRT の推定結果を利用した項目精選の結果から、選ばれた項目や選ばれなかった項目を示すことができたことから、今後も調査に必要な項目であるのか、または検討する必要がある項目なのかを示すことができた。以上のことから、項目の良し悪しを示すことができたことにより、項目自体の評価も行える可能性を示すことができた。

今後の課題としては、項目選択の方法として、変数減少法を用いたが、真の解としての総当たり法との比較を行う必要がある。また、今回は次元性が確認できたデータのみを利用したが、カテゴリカルデータは多次元性をもったデータも多く存在するため、これらのデータに対する検討も課題として残されている。

謝辞

本研究において、岡山理科大学 経営学部 経営学科 教授 森 裕一 先生、同学部学科 教授 黒田 正博 先生をはじめ、諸先生方にご指導を賜り、論文にまとめることができました。ここに深い感謝の意を表します。

また、「東京大学社会科学研究所人材ビジネス研究寄付研究部門」よりデータをご提供頂き、有益な情報を得ることができました。心より御礼申し上げます。

参考文献

- 加藤健太郎, 山田剛史, 川端一光 (2014). R による項目反応理論. オーム社.
- 並川努, 谷伊織, 脇田貴文, 熊谷龍一, 中根愛, 野口裕之 (2012). Big Five 尺度短縮版の開発と信頼性と妥当性の検討. 心理学研究, 第 83 巻, 2 号, 91-99.
- 並川努(2015). 心理尺度短縮版における IRT の活用に関する研究. 名古屋大学, 乙第 7122 号, 2-108.
- samejima, F. (1969). A general model for free-response data. *Psychometric Monograph*, 17.
- 笹川智子, 金井喜宏, 村中泰子, 鈴木伸一, 嶋田洋徳, 坂野雄二 (2004). 他社からの否定的

評価に対する社会的不安測定尺度（FNE）短縮版作成の試み - 項目反応理論による検討
- 行動療法研究, 第 30 巻第 2 号, 87-98.

東京大学社会科学研究所人材ビジネス研究寄付研究部門 (2004). 在宅介護ヘルパーの仕事と
能力に関するアンケート.

東京大学社会科学研究所人材ビジネス研究寄付研究部門 (1998). 職業能力評価および資格の
役割に関する調査.

浦上昌則, 脇田貴文 (2016). 項目反応理論を用いた進路選択に対する自己効力尺度短縮化の
試み. 南山大学紀要「アカデミア」人文・自然科学編, 第 12 号, 67-76.

脇田貴文 (2004). 評定尺度法におけるカテゴリ間の間隔について-項目反応モデルを用いた
評価方法—. *The Japanese Journal of Psychology* 2004, Vol.75, No4, 331-338.