

ベイズ IRT による間接的発話要求の新たな理解

大石 信弘^{1,*}

New Understanding of Indirect Speech Requests by Means of Bayesian Item Response Theory

Nobuhiro Oishi^{1,*}

A questionnaire survey on indirect speech requests conducted by Hirakawa was analyzed using Item Response Theory (IRT) applying Bayesian statistical modeling. From the knowledge of both the characteristics of the respondents and the characteristics of the question items, it is found that the characteristics of the question items are the cause of the bimodal distribution of respondents' tendency to interpret the "interpretation of the request". Also Bayesian IRT appears to offer benefits for scale design.

キーワード：間接的発話要求、項目反応理論、MCMC、ベイズ統計モデリング、ベイズ IRT

Keywords : Indirect Speech Request, Item Response Theory, Markov Chain Monte Carlo Method, Bayesian Statistical Modeling, Bayesian IRT

1. はじめに

近年のデータサイエンスの進展には目を見張るものがあり、いろいろな場面においてデータに基づいた言説が重視されるようになってきた。データの分析を主眼にした数理モデルの適用分野を数え上げるだけでも、統計学、物理学、化学、生物学、エコロジー、医学、心理学、教育学、経済学など、理工系に限らず文系をも含めた広い範囲に渡っている。

特に心理学の分野では、個性の有様を組み込むような柔軟なモデル構築が可能な、ベイズ統計を用いた研究が盛んにおこなわれるようになった⁽¹⁾。

ここでは一例として、平川等による間接的発話要求に関する研究⁽²⁾⁽³⁾⁽⁴⁾を取り上げる。間接的発話要求とは、例えば、「この部屋暑いね」と発話することで、遠回しに窓を開けるよう要求するような、遠回しな言い方による他者への依頼である。平川は、間接的発話要求が発話される目標を解明して、それに沿った分類を行い⁽³⁾⁽⁴⁾、さらにベイズモデリングにより、間接的発話要求の受け手には、要求を妥当と解釈しやすいグループとそうでないグループの2つのグループが存在することを明らかにした⁽²⁾。

一方、教育学の分野における数理モデルの適用例とし

て、古典的なテスト理論に代わる項目反応理論 (IRT: Item Response Theory)⁽⁵⁾が挙げられる。IRTでは、項目(出題問題)の特性と受験者の特性とを区別して、テストの結果を評価するため、異なる集団が同じテストを受験しても、統一的に受験者を評価することができるようになる。

本研究では、平川が明らかにした、間接的発話要求の受け手が2つのグループに分かれることを、IRTにより分析することで、発話項目と関連付けた理解を試みる。

筆者はこれまで、統計解析環境 R⁽⁶⁾を用いて、因果分析⁽⁷⁾や教育ツールの開発⁽⁸⁾を行ってきた。また、時期と受験者の異なる小規模集団を対象とした IRT に、ベイズモデリングを適用してきた⁽⁹⁾。本研究ではこれらに倣い、統計統合環境 R ver. 4.2.2 を用いた。特に、ベイズ統計モデリングに用いるマルコフ連鎖モンテカルロ法 (MCMC: Markov Chain Monte Carlo) には rstan パッケージ(2.26.17)を用いた。PC は Core i9-12900U CPU(16 コア、24 スレッド)、32GB RAM の Windows 11 Pro(64bits)22H2 である。

2. 平川による先行研究

2.1 分析に用いるデータ

平川は、間接的発話要求の受け手の特性を明らかにするために、「要求」に絞った「間接的発話」を、「要求と解釈するかどうか」について 400 名を対象にアンケート調査し、結果とともに調査データを公開している⁽²⁾。アンケ

¹ 電子情報システム工学系
〒861-1102 熊本県合志市須屋 2659-2
Faculty of Electronics and Information Systems Engineering,
2659-2 Suya, Koshi-shi, Kumamoto, Japan 861-1102
²
* Corresponding author:
E-mail address: oishi@kumamoto-nct.ac.jp (N. Oishi).

表 1 間接的発話要求の質問項目

質問番号	場面と発話	要求の解釈	解釈を妥当と回答した数	解釈を不適当と回答した数
p1	あなたは友人の A さんと 2 人で窓の閉まった部屋にいます。その時、A さんが「この部屋暑いね」と言いました。	A さんはあなたに、窓を開けるよう頼んでいる。	153	247
p2	あなたは友人の A さんと 2 人でいて、A さんはいくつかの荷物を運んでいます。その時、A さんが「重いな」と言いました。	A さんはあなたに、荷物を持つのを手伝うよう頼んでいる。	200	200
p3	あなたは友人の A さんと一緒に 1 つの大きな荷物を運んでいます。その時、A さんはあなたに「ちょっと速いよ」と言いました。	A さんはあなたに、もっとゆっくり荷物を運ぶよう頼んでいる。	375	25
p4	あなたは友人の A さんと 2 人でいて、A さんは何か作業をしています。その時、A さんが「この作業、大変だなあ」と言いました。	A さんはあなたに、作業を手伝うよう頼んでいる。	142	258
p5	あなたは友人の A さんとテレビを見ています。その時、A さんが「ちょっと聞こえにくいな」と言いました。	A さんはあなたに、テレビの音量を大きくするよう頼んでいる。	272	128
p6	あなたはとあるセミナーに出ています。その時、隣の席に座る友人の A さんが「消しゴムある？」と言いました。	A さんはあなたに、消しゴムを貸すよう頼んでいる。	362	38
p7	あなたは友人の A さんと 2 人で食事にレストランに行きました。食事後、A さんが「あ、財布わすれちゃった」と言いました。	A さんはあなたに、お金を貸すよう頼んでいる。	210	190

表 2 アンケート回答のマイクロデータ (抜粋)

id	性別	年齢	質問項目への回答							妥当とした数 y
			p1	p2	p3	p4	p5	p6	p7	
1	F	37	0	0	1	0	0	1	1	3
2	F	32	1	1	1	1	0	1	1	6
3	M	59	1	0	1	0	1	1	1	5
4	F	33	0	1	0	0	1	1	1	4
5	F	55	1	1	1	1	1	1	1	7
~										
400	M	51	0	0	1	0	0	1	0	2

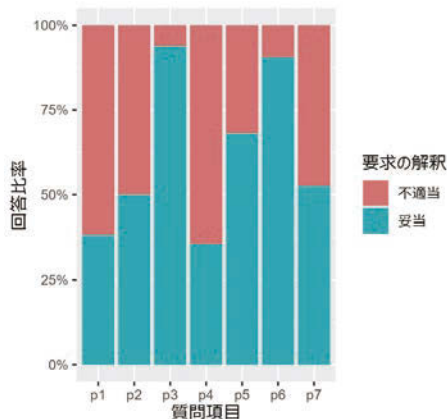


図 1 各質問項目の回答割合

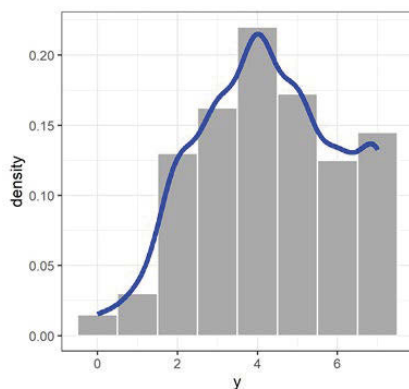


図 2 妥当とした数 y の 2 峰性分布

ートの実施は Web 調査会社が行い、回答者は 17~59 歳のアンケートモニタ 400 名 (男性 199 名、女性 201 名、平均年齢 41.5 歳) であった。回答者に提示されたアンケート項目を表 1 に示す。7 つの間接的発話要求の質問項目 (p1~p7) が「場面と発話文」として示され、それに対する「要求の解釈」を明示し、そう解釈するのが妥当か妥当でない (不適当) かを回答してもらっている。発話する際の詳細な状況を設定していないため、質問項目に対する正解はなく、あくまでも個人の主観による判断で回答されているのが特徴である。アンケートの結果は表 2 のようなマイクロデータとして公開され、研究に供されている⁽²⁾。表中、p1~p7 には、各項目に対応する要求の解釈が妥当であれば 1、妥当でない (不適当) であれば 0 の 2 値データが入っている。また、表中の y は、各回答者が質問項目に妥当と回答した合計であり、各人が解釈を妥当としやすい傾向を表している。各質問項目において妥当と解釈された比率を図 1 に示す。p2 と p7 はどちらも、妥当と不適当の割合がほぼ半々であり、それに対して、p1 と p4 はどちらも不適当と回答した割合が 60 数% と多数派である。これとは逆に、p3 と p6 はどちらも妥当と回答した割合が 90% を超える多数派である。図 2 に、妥当とした数 y のヒストグラムと密度関数 (青線) を示す。y の平均値は 4.3 であり、ヒストグラムの最頻値 4 とほぼ一致している。密度関数を見ると、最頻値に大きなピークが認められるものの、y=7 付近にもう一つのピークを持つ 2 峰性の分布になっていることが分かる。

2.2 平川モデル

平川は、妥当とした数 y を生成するモデルとして、p1~p7 の各質問項目の「要求の解釈」を妥当とみなす確率 (以後「解釈率」と略記する) が、個人ごとの特有の値によって決まる 2 項分布モデルを採用した。試行回数は項目数と同じ T=7 である。これを踏まえて、図 2 の 2 つの

ピークの由来として、解釈を不適当とみなしやすいグループ A (解釈率 θ_1) と、解釈を妥当とみなしやすいグループ B (解釈率 θ_2) の存在を想定した。回答者がグループ A に所属する確率を π (定数) として、 y はこの 2 つのグループの重ね合わせ、つまり、混合 2 項分布によって生起するとした。

$$\begin{aligned} & Binomial_Mix(y|\pi, T, \theta_1, \theta_2) \\ &= \pi \times Binomial(y|T, \theta_1) \\ &+ (1 - \pi) \times Binomial(y|T, \theta_2) \\ &y_j \sim Binomial_Mix(\pi, 7, \theta_1, \theta_2) \dots\dots\dots (1) \end{aligned}$$

このモデルを平川モデルと呼ぶことにする。MCMC による事後推定値を EAP(post. sd)[95%確信区間] の書式で記すと、 π は、0.743(0.045)[0.648, 0.823]、 θ_1 は 0.509(0.017)[0.474, 0.542]、 θ_2 は 0.913(0.026)[0.859, 0.960] と報告されている。つまり、解釈の是非を半々と考える人が全体のおよそ 3/4 いて、残りの 1/4 の人は、要求を 91.3% という高い確率で妥当と解釈するという、2 つのグループに分けられると結論付けている。

3. ベイズ IRT

平川が行ったアンケートを、発話項目の特性と関連付けて解釈するために、次式で示す IRT の 2PLM モデル (2 Parameters Logistic Model) を適用する。

$$P_{ij}(a_j, b_j, \theta_i) = \frac{1}{1 + \exp\{-1.7a_j(\theta_i - b_j)\}} \dots (2)$$

ここで、 a_j は識別力パラメータ、 b_j は困難度パラメータであり、 j 番目の発話項目を特徴づけるパラメータである。 θ_i は、 i 番目の回答者を特徴づけるパラメータであり、ここでは回答者の解釈傾向と呼ぶことにする。 P_{ij} は、 i 番目の回答者が j 番目の発話項目に対して「要求の解釈」を妥当とみなす確率である。この確率が 0.5 以上であれば「解釈は妥当」という回答が得られ、0.5 未満であれば「解釈は不適当」が得られるとする。

このようにして得られた回答パターンを \mathbf{u} として、パラメータ $\boldsymbol{\lambda}$ を推定するためにベイズの定理を適用する。回答パターンが与えられた時のパラメータの分布 $f(\boldsymbol{\lambda}|\mathbf{u})$ は、ベイズの定理より次式で与えられる。

$$f(\boldsymbol{\lambda}|\mathbf{u}) \propto Y(\mathbf{u}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}) \dots\dots\dots (3)$$

ここで、左辺 $f(\boldsymbol{\lambda}|\mathbf{u})$ は事後分布、右辺の $Y(\mathbf{u}|\boldsymbol{\lambda})$ は、パラメータを指定したもとの回答パターンの尤度分布、 $p(\boldsymbol{\lambda})$ はパラメータの事前分布である。本研究では、識別力パラメータおよび困難度パラメータの事前分布として、正規分布を採用した。また、(2) 式の尤度関数としてベルヌイロジット関数を用いた。以下にこれらの尤度関数を示す。

$$a_j \sim Norm(0, 2) \quad j = 1, \dots, 7$$

$$b_j \sim Norm(0, 2) \quad j = 1, \dots, 7$$

$$Y_j \sim BernoulliLogit\{1.7a_j(\theta_i - b_j)\} \quad i = 1, \dots,$$

$$400; j = 1, \dots, 7 \dots\dots\dots (4)$$

バーンイン期間 $B = 10^4$ 、有効乱数発生数 $M = 9 \times 10^4$ の乱数列を 4 チェーン発生させ、rstan パッケージを用いて MCMC を実行した。トレースプロットや自己相関関数、並びに収束判定指標 \hat{R} から、発生した乱数は十分収束しているとみなせた。

4. 分析結果および考察

4.1 各パラメータの事後分布

ベイズ IRT で求めた、各パラメータの事後分布を表 3 に示す。第 1 列目は、推定した各パラメータ名、つまり、各回答者の解釈傾向 $\theta_i (i = 1 \sim 400)$ 、各項目の識別力パラメータ $a_j (j = 1 \sim 7)$ 、および困難度パラメータ $b_j (j = 1 \sim 7)$ である。第 2 列目は事後期待値 EAP (Expected A Posteriori)、3 列目は事後標準偏差、4 列目以降は 95% 確信区間、ならびに、四分位範囲である。各パラメータの点推定量としては、いくつかの流儀があるが、ここでは EAP を採用する。 θ_i の最小値、最大値および EAP の平均値は、それぞれ、-1.98、1.33 および 0.00 であった。 a_j は非負であり、その値が小さいと、「要求の解釈」を妥当と解釈しやすい傾向の人かそうでないかを判別しにくくなる。今回の 7 つの項目だと、 a_6 と a_7 の値が他と比べて小さくなっている。 b_j は、-2.91 から 0.50 の値を取った。各項目の回答における、妥当か不適当かの割合を反映している。

4.2 解釈傾向について

回答者の解釈傾向 θ_i の代表値である EAP のヒストグ

表 3 各パラメータの事後分布

パラメータ	EAP	post. sd	2.5%	25%	50%	75%	97.5%
θ_1	-0.749	0.559	-1.908	-1.109	-0.726	-0.364	0.288
~							
θ_{400}	-1.081	0.596	-2.318	-1.469	-1.055	-0.667	0.016
a_1	1.962	0.336	1.380	1.717	1.922	2.154	2.693
a_2	2.156	0.389	1.508	1.884	2.116	2.385	3.030
a_3	1.303	0.312	0.765	1.082	1.278	1.496	1.985
a_4	1.921	0.331	1.355	1.689	1.893	2.121	2.650
a_5	1.480	0.256	1.027	1.302	1.463	1.639	2.032
a_6	0.920	0.215	0.545	0.767	0.905	1.055	1.383
a_7	1.018	0.182	0.683	0.891	1.009	1.136	1.398
b_1	0.402	0.092	0.228	0.339	0.399	0.461	0.590
b_2	0.000	0.081	-0.159	-0.055	0.000	0.054	0.159
b_3	-2.694	0.498	-3.898	-2.950	-2.614	-2.346	-1.961
b_4	0.504	0.097	0.323	0.437	0.500	0.566	0.704
b_5	-0.719	0.125	-0.990	-0.797	-0.711	-0.633	-0.497
b_6	-2.909	0.597	-4.342	-3.229	-2.814	-2.485	-2.019
b_7	-0.120	0.125	-0.372	-0.201	-0.118	-0.037	0.121

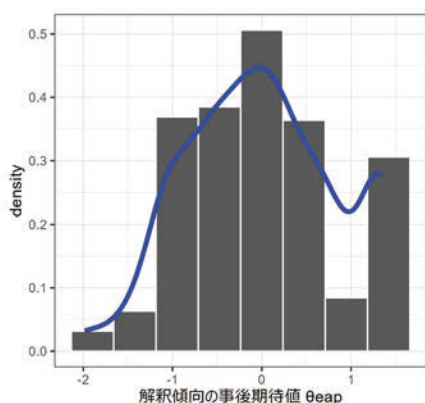


図3 解釈傾向の事後期待値のヒストグラム

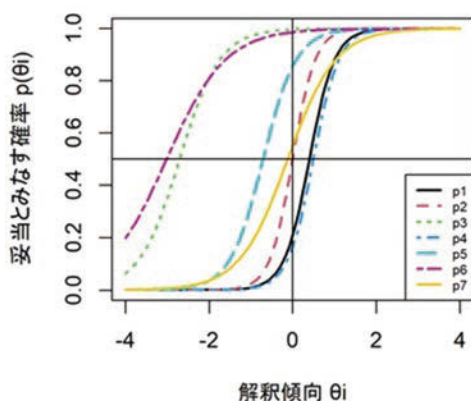


図4 項目特性曲線

ラムを図3に示す。ヒストグラムの階級数は、図2のヒストグラムのそれと同じ8つにしている。図中の青線は密度関数を示している。図2の2峰性の密度関数と同じように、一番右の階級にもピークを持つ2峰性を示している。

4.3 質問項目の特徴

IRTの特徴である質問項目についての知見を得るために、各項目の識別力パラメータ a_j と困難度パラメータ b_j のEAPを用いて描いた項目特性曲線(ICC: Item Characteristic Curve)を図4に示す。

ここで、図1の回答割合が同程度だった、(p1とp4)および(p2とp7)の組合せに着目する。

p1とp4のICCはほぼ重なっている。これは、「解釈を妥当」と回答した割合ばかりでなく、p1とp4で同じ回答をした回答者が、ほぼ全員であったことを示す。最適なアンケート項目の観点からは、冗長であるためどちらかの質問項目は省略できると思われる。不適当とした割合が60数%であったため、グラフは右側に現れている。

p2とp7では、妥当とした割合が50%程度であり、 $b_2 \approx b_7 \approx 0$ となっている。ただ、 $a_7 < a_2$ であり、両者のICC

は一致せず、p7の曲線が緩やかに上昇しているため、 $\theta_i = 1$ 付近でも確率が1に達しておらず、p1やp4と逆転している。 θ_i が2を過ぎて、やっと全ての項目の確率が1となる。この逆転現象が、解釈傾向のヒストグラムの2峰性の原因であると考えられる。

なお、項目情報曲線およびテスト情報曲線を用いた吟味については、紙面の都合上割愛する。

5. まとめ

今回、平川が行った間接的発話要求に関するアンケート調査に、ベイズIRTを適用することで、回答者の特徴と質問項目の特徴の両方を明らかにした。それらの知見により、間接的発話要求の解釈傾向の分布が2峰性を示すことを、質問項目の特徴と関連付けて理解できることが分かった。

また、質問項目の特徴を明らかにすることで、アンケート調査に用いる質問項目を吟味できることが分かった。このことは、回答者の特徴を抽出するための、よりよい質問項目の構成に役立つものと考えられる。

さらに、ベイズ統計モデリングによるIRTを採用したため、サンプルサイズの小さい小規模集団を対象とした調査が可能となり、尺度構成デザインにメリットをもたらすと思われる。

(令和5年9月8日受付)

(令和5年10月30日受理)

参考文献

- (1) 豊田秀樹:「基礎からのベイズ統計学」, pp.i-vi, 朝倉書店(2015).
- (2) 豊田秀樹, 平川真 他:「たのしいベイズモデリング」, 第4章, pp.33-42, 北大路書房(2018).
- (3) 平川真:「自己—他者配慮的目標が間接的発話要求の使用に及ぼす影響」, 心理学研究, 82巻, pp.532-539(2012).
- (4) 平川真, 森永康子:「間接的発話要求によって使用者の目標は達成できるか」, 対人コミュニケーション研究, 第2号, pp.19-30(2014).
- (5) 加藤健太郎, 山田剛史, 川端一光:「Rによる項目反応理論」, pp.70-281, オーム社(2014).
- (6) The R Foundation: <https://www.r-project.org/foundation/>, Retrieved Sep. 8, 2023.
- (7) N. Oishi, N. Yamamoto, A. Ishida and J. Murakami: “A Causal Analysis by Structural Equation Modeling of Sleep Monitoring Sensor Data”, IJEEE, vol. 8, no. 3, pp. 58-62, 2020.
- (8) 石田明男, 扇崎和希, 山本直樹, 大石信弘, 村上純: 「第7号60頁の4×4×4立方体パズルについて(後編)」, 初等数学, 第90号, pp.18-22(2021).
- (9) 大石信弘:「ベイズIRTと水平等化を用いた小規模テストの年度間比較」, 熊本高等専門学校 研究紀要, 第14号, pp.24-27(2023).