

令和 3 年 6 月 22 日現在

機関番号：12102

研究種目：基盤研究(C) (一般)

研究期間：2018～2020

課題番号：18K03057

研究課題名(和文) 誤答理由を推定する認知診断モデル開発とその実践的適用

研究課題名(英文) Development of cognitive diagnostic model to estimate the reason for wrong answers and its practical application

研究代表者

尾崎 幸謙 (Ozaki, Koken)

筑波大学・ビジネスサイエンス系・准教授

研究者番号：50574612

交付決定額(研究期間全体)：(直接経費) 2,000,000円

研究成果の概要(和文)：本研究は能力測定のための多肢選択型テストにおいて誤答した受検者の誤答理由を受検データから推定するための統計モデルを考案し、そのモデルを実際のデータに適用して、実用可能性を検討することであった。3年間の研究成果として、Ozaki, Sugawara, & Arai (2020)において統計モデルを発表することができた。また、上記論文においてリーディングスキルテストへの適用を行い、「問われている語と最も近い位置にある語が主語である」「文頭の語が必ず主語である」という2つの誤答理由を抽出することに成功した。また、さらなる適用可能性を検討するために、データリテラシーを測定するテストデータを収集した。

研究成果の学術的意義や社会的意義

テストの受検結果を次の学習に活かすためには、点数や偏差値だけでなく、各受検者の誤答理由を明らかにすることが重要である。誤答理由が明らかになれば、指導者は今後の指導方針を立てやすくなり、学習者は復習すべき箇所が明確になるからである。本研究は、多肢選択型のテスト問題に対する回答データから、各受検者がどのような誤答理由に基づいて誤答してしまったかを新たに開発した統計モデルに基づいて明らかにすることを目指したものである。開発した確率モデルを読解力調査やデータリテラシー調査に適用し、その実用可能性を検討した。

研究成果の概要(英文)：The purposes of this study were to develop a statistical model for estimating the reason for the wrong answer of the examinee who answered wrongly in a multiple-choice test for measuring ability from the test data, and to apply the model to the actual data to examine the practicality. As a result of three years of research, the statistical model was shown in Ozaki, Sugawara, & Arai (2020). In addition, in the above paper, the model was applied to the reading skill test and extracted two reasons for incorrect answers: "the word closest to the questioned word is the subject" and "the first word is always the subject". Test data to measure data literacy to examine further applicability was also collected.

研究分野：テスト理論

キーワード：認知診断モデル 読解力 データリテラシー 誤答分析

1. 研究開始当初の背景

能力測定データの分析方法として一般的なものは、合計点や偏差値を求める方法である。一方で、近年では項目反応理論というテスト理論に基づいた能力評価が行われることもある。これらの方法に共通しているのは、測定対象となっている構成概念の値の高低を一次元上の数値で表したものであるということである。これらの数値は能力の高低を表しているという意味で、受検者個人の学習成果の確認や合否判定等に利用することができるが、テストデータから知り得るものはこれだけではない。

テスト理論の分野で近年注目を集めている統計モデルに、認知診断モデルというものがある。これは、問題を解くために必要な能力を細分化して、細分化された個々の能力を各受検者が持っているか否かを推定することが可能な統計モデルである。たとえば、分数の足し算の問題であれば、通分する能力・(通分した後に)分子だけを足す能力などが考えられる。認知診断モデルではテストに含まれる各問題がどのような(細分化された)能力を測定するものであるかを予め設定しておき、各問題に対する正誤データから、個々の能力測定を行うものである。細分化された能力を測定することで、受検者にとっては今後の学習計画を細かく立てることが可能となり、指導者にとってはきめの細かな学習指導が可能となる。

一方で、問題に誤答したときにその理由を知ることが教育上は非常に重要である。誤答した理由を知ることができれば、理解の誤り等の改善に直接的に役立つからである。正答するために必要な能力の有無(正答できた理由)を測定する統計モデルに関しては盛んに研究されていた一方で、誤答した理由を推定する統計モデルの研究は Kuo et al. (2016, 2018)が挙げられる程度であった。

2. 研究の目的

本研究の目的は以下の2つであった。1つめは、多肢選択型の問題に対する受験者の回答データから、受検者が誤答した理由を推定するための統計モデルを開発することであった。2つめは、開発したモデルを実際のテストデータに適用し、この方法の実用可能性を検討することであった。Kuo et al. (2016, 2018)は多肢選択型の問題に対する回答データを正誤の2値に変換した後に、DINOモデルという認知診断モデルを適用することで誤答理由を推定することができることを明らかにした。彼らはこのモデルを Bug-DINO モデルと呼んでいる。

しかしながら、元々多肢選択型の問題に対する回答データを正誤の2値に変換するというのは情報を欠落させてしまうことになるため、誤答理由の推定精度が下がってしまう可能性がある。このため、本研究では多肢選択型の問題に対する回答データをそのまま扱う統計モデルの開発を目的とした。

また、認知診断モデルは、問題に正答するために必要なスキルを予め漏れなく設定する必要があり、このことが実用上の難しさを生んでいる。誤答理由についても各問題について予め漏れなく設定することが望ましいが、網羅していなかったとしても設定した誤答理由についてはその有無を推定できると考えるため、実用上の困難はそこまで大きくはないと想定していた。しかし、その想定の上しさ等を調べるために、2つめの目的である実データへの適用を行った。

3. 研究の方法

(1) モデル開発

多肢選択型項目に対する認知診断モデルとして研究代表者が開発した Ozaki (2015)のモデルをベースとした。このモデルは多肢選択型項目に対する回答データから、問題を解くために必要な個々の能力を受検者が有しているか否かを推定するためのモデルである。研究代表者らはこのモデルをベースとして、多肢選択型項目に対する回答データから、受検者の誤答理由を推定するモデルを考案し、Ozaki, Sugawara, & Arai (2020)で発表した。考案したモデルは2つあり、それぞれに MC-DINO1, MC-DINO2 と命名した。2つのモデルの違いはモデルの複雑さ(パラメータ数の多さ)である。実際のデータに対して適用する際には、情報量規準等を用いて適当なモデルを判断する。

Ozaki, Sugawara, & Arai (2020)では開発した2つのモデルと既存の Bug-DINO モデルとの比較をシミュレーション研究によって行っている。シミュレーション研究では、5つの誤答理由を定め、各5つを受検者が有しているか否かを予め設定しておいた。そして、モデルのパラメータ等を定めたうえで、4肢選択型項目に対する回答データを(MC-DINO1モデルまたはMC-DINO2モデルに基づいて)発生させた。その回答データを Bug-DINO モデル, MC-DINO1 モデル, MC-DINO2 モデルのそれぞれで分析して、予め定めた各受検者の誤答理由保有状況とどの程度一致するかを確認した。

(2) 実データへの適用

実際のテストとして、一般社団法人教育のための科学研究所が実施する読解力テストであるリーディングスキルテスト (Arai, et al 2017) を選んだ。また、数値や言語情報に基づいて客観的に正しい判断を行う能力を測定するためのデータリテラシーテストを研究代表者が開発し、その受検データに対しても適用を行った。

4. 研究成果

(1) シミュレーション研究の結果

開発された MC-DINO1 モデルと MC-DINO2 モデルの詳細については複雑な数式になるので Ozaki, Sugawara, & Arai (2020) を参照されたい。

下記の表 1 から表 3 はシミュレーション研究の結果の一部である。表 1 は項目数 (問題数) が 10、受検者数が 250 人の場合の結果である。左は誤答理由の保有率が 20%、右は 40% の場合である。真のモデルとは、4 肢選択型項目に対する回答データを発生させるときのモデルを表しており、MC1 は MC-DINO1 モデル、MC2 は MC-DINO2 モデルのことである。表 2 は受検者数を 500 人にした場合、表 3 は項目数を 20、受検者数を 500 人にした場合を表している。

誤答理由の保有率20%				誤答理由の保有率40%			
真のモデル	Bug-DINO	MC1	MC2	真のモデル	Bug-DINO	MC1	MC2
MC1	0.275	0.970	0.955	MC1	0.412	0.972	0.970
MC2	0.303	0.879	0.855	MC2	0.444	0.892	0.905

誤答理由の保有率20%				誤答理由の保有率40%			
真のモデル	Bug-DINO	MC1	MC2	真のモデル	Bug-DINO	MC1	MC2
MC1	0.507	0.983	0.974	MC1	0.521	0.978	0.976
MC2	0.357	0.878	0.854	MC2	0.465	0.891	0.905

誤答理由の保有率20%				誤答理由の保有率40%			
真のモデル	Bug-DINO	MC1	MC2	真のモデル	Bug-DINO	MC1	MC2
MC1	0.623	0.999	0.999	MC1	0.751	0.995	0.995
MC2	0.611	0.969	0.981	MC2	0.662	0.953	0.969

表 1 から表 3 から分かるように、既存の Bug-DINO モデルと比べて、開発した 2 つのモデルの推定精度は極めた高かった。特に、受検者数や項目数が少ない場合であっても 90% 程度の精度を持つことが示された。

(2) リーディングスキルテストへの適用結果

リーディングスキルテスト (Arai, et al 2017) は、一般社団法人教育のための科学研究所が実施する読解力テストであり、読解力の 6 つの側面を項目反応理論に基づいて測定することが可能である。リーディングスキルテストは、誤答理由を推定することを目的としたものではないが、元々適用を念頭に置いていないテストに適用することで開発したモデルの実用可能性が上がると考え、適用するデータとした。誤答理由としては「問われている語と最も近い位置にある語が主語である」「文頭の語が必ず主語である」の 2 つを予め定めた。

リーディングスキルテストの一例を図 1 に示した。この問題の正解はキリスト教であるが、イスラム教を選んだ場合には「問われている語と最も近い位置にある語が主語である」という誤った考え方を有している可能性が高く、仏教を選んだ場合には「文頭の語が必ず主語である」という誤った考え方を有している可能性が高いと考えた。

以下の文を読みなさい。

仏教は東南アジア，東アジアに，キリスト教はヨーロッパ，南北アメリカ，オセアニアに，イスラム教は北アフリカ，西アジア，中央アジア，東南アジアにおもに広がっている。

この文脈において，以下の文中の空欄にあてはまる最も適当なものを選択肢のうちから1つ選びなさい。

オセアニアに広がっているのは（ ）である。

- キリスト教 イスラム教
- ヒन्दウー教 仏教

図1 . リーディングスキルテストの問題例 (<https://www.s4e.jp/example#!#carousel-rst> より転載)

このような問題を8問用意し，リーディングスキルテストで収集された281人の受検者データ(40人の中学生と241人の高校生)を分析したところ，「問われている語と最も近い位置にある語が主語である」の保有率は2.7%，「文頭の語が必ず主語である」の保有率は3.0%であると推定された。この保有率は値としては低い，このような生徒の存在を明らかにできたことは教育上有効な示唆を与えるものである。

(3) データリテラシーテストへの適用への適用結果

研究代表者は下記のようなデータリテラシーを測定するテストを作成した。このテストは，数値や言語情報に基づいて客観的に正しい判断を行う能力を測定するためのものであった。

以下の2つの図は，その日のビールの売上と前日のビールの売り上げの関係(左図)および，その日のビールの売り上げとある放送局の前日のビールに関するCM放映回数(右図)を示したものである。売り上げの単位は円である。データはある月の30日間についてのものである。2つの図から読み取ることができることとして最も適切なものを選びなさい。

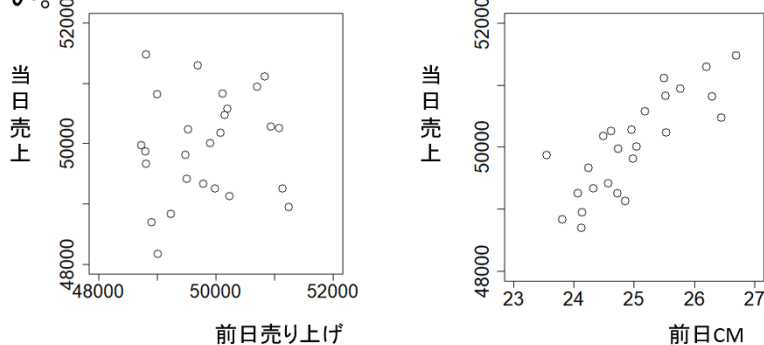


図2 . データリテラシーテストの問題例

上記の問題の選択肢は以下であった。

1. ビールは冬よりも夏の方がよく売れる。
2. 平日よりも土日の方がビールの売り上げは高い。
3. ビールの売り上げには，CM放映回数よりも前日のビールの売り上げの方が関係している。
4. 前日のビールの売り上げがいくらであったとしても，当日のビールの売り上げはおおよそ48000円から52000円の間である。

正解は4である。1や2の選択肢は問題文中で全く触れられていない内容ではあるが，特に1は常識的には正しそうな内容である。このような内容の選択肢を選ぶ場合には，データに基づいた理解ではなく主観的な判断を行っていることになる。そこで，このような「書かれていないことに基づいて判断してしまう誤り(主観的判断をしてしまう誤り)」などの誤った考え方を定め，リーディングスキルテストと同じような分析を遂行中である。

(4) 研究成果の社会還元・今後の展望

誤答理由を推定することは、学習者の今後の学習方針や指導者の今後の指導方針を定めるために極めて有効である。そのための方法論を開発できたことは、社会に対して一定の貢献ができたものとする。

今後の展望としては、多肢選択型問題に対する回答データから、問題に正答するために必要なスキルと、誤答理由の双方を同時に推定するモデルの開発である。学習者や指導者が知りたいことは、どちらか一方ではなく、その双方である。このモデルを開発することで、学習者や指導者にとってより有益な情報提供が可能と考える。

<引用文献>

Arai, NH, Todo, N, Arai, T, Bunji, K, Sugawara, S, Inuzuka, M, Matsuzaki, T, Ozaki, K (2017) Reading Skill Test to Diagnose Basic Language Skills in Comparison to Machines. Proceedings of the 39th Annual Cognitive Science Society Meeting (CogSci 2017): 1556-1561.

Kuo B-C, Chen C-H, Yang C-W, Mok MMC (2016) Cognitive diagnosis models for test with multiple-choice and constructed-response items. Educational Psychology 36(6): 1115-1133.

Kuo B-C, Chen C-H, de la Torre J (2018) A cognitive diagnosis model for identifying coexisting skills and misconceptions. Applied Psychological Measurement 42(3): 179-191.

Ozaki K (2015) DINA models for multiple-choice items with few parameters: considering incorrect answers. Applied Psychological Measurement 39(6): 431-447.

Ozaki K, Sugawara S, Arai, N (2020) Cognitive diagnosis models for estimation of misconceptions analyzing multiple-choice data. Behaviormetrika 47(1): 19-41.

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Ozaki, K., Sugawara, S. & Arai, N.	4. 巻 47
2. 論文標題 Cognitive diagnosis models for estimation of misconceptions analyzing multiple-choice data.	5. 発行年 2020年
3. 雑誌名 Behaviormetrika	6. 最初と最後の頁 19-41
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s41237-019-00100-9	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計5件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 児島健、石澤洋子、尾崎幸謙、山形伸二
2. 発表標題 批判的思考力を支えるリテラシーテスト開発の試み1 - 論証リテラシー -
3. 学会等名 産業・組織心理学会第34回全国大会
4. 発表年 2018年

1. 発表者名 石澤洋子、児島健、山形伸二、尾崎幸謙
2. 発表標題 批判的思考力を支えるリテラシーテスト開発の試み2 - データリテラシー -
3. 学会等名 産業・組織心理学会第34回全国大会
4. 発表年 2018年

1. 発表者名 尾崎幸謙
2. 発表標題 複数概念を測定するための適用型テストの検討
3. 学会等名 日本行動計量学会第46回大会
4. 発表年 2018年

1. 発表者名 登藤直弥、分寺杏介、尾崎幸謙
2. 発表標題 出題領域と解答時間を考慮した項目選択アルゴリズムの性能評価
3. 学会等名 日本行動計量学会第46回大会
4. 発表年 2018年

1. 発表者名 歌川真一郎、尾崎幸謙
2. 発表標題 複数の能力を測定する適用型テストにおける効率化の方法
3. 学会等名 日本テスト学会第16回大会
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担者	菅原 真悟 (Sugawara Shingo) (00745052)	国立情報学研究所・社会共有知研究センター・特任研究員 (62615)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------