

確率論に関する類似問題と難易度の関係性の分析

阿部 佳宣[†] 高木 正則[‡] 市川 尚[‡]

岩手県立大学 ソフトウェア情報学部[†]

1 はじめに

コンピュータを用いて行うテスト手法の一つに、コンピュータ適応型テスト (Computerized adaptive Testing, 以下, CAT) がある[1]. CAT は学習者の回答状況から逐次学習者の能力を推定し, 学習者に適した問題をコンピュータが判断して提示し, 回答させるテスト手法であり, 短時間で高精度に能力値を推定できるというメリットがある. しかし, CAT を行うためには予め問題とその問題の特性値を格納するデータベース (以下, アイテムバンク) が必要であり, この特性値を求めるためには, 一般的に項目反応理論 (Item Response Theory, 以下, IRT) が用いられているが[2][3], IRTの分析を行うためには膨大な回答データが必要であるという問題点がある. この問題に対し, 我々は特性値が既知の問題から問題文, 前提条件, 問い方, 問題文中の数値の4つの条件の中で一致する条件が1つ以上ある, かつ異なる条件が1つ以上ある問題を抽出しその類似問題の類似度や差分情報に基づいて特性値を推定する手法を検討してきた. 本研究では, 数学の確率論に関する類似問題と難易度の関係性を明らかにすることを目的とし, 類似問題間で困難度に差が生じた問題と, 困難度に差が生じなかった問題を比較し, 類似問題間の差異が項目特性に与える影響について考察する.

2 IRT

CATを行うために必要な情報として項目特性と呼ばれる特性値がある. この項目特性には能力の高い回答者と低い回答者をどの程度区別できるかを表す指標である識別力と問題の難易度を表す指標である困難度等がある.

この項目特性を求めるためには一般的にIRTが用いられている. IRTでは困難度のみを求める1パラメータロジスティックモデルや識別力と困難度を求める2パラメータロジスティックモデルなどがあるが, 項目特性を求めるためには回答データ数が前者で200件以上, 後者で500件以上必要である[4]. また, パラメータを推定後, CATに用いるのに適切な困難度と識別力の問題を選別するため, アイテムバンクの構築には多大なコストがかかるという問題がある.

Analysis of the relationship between problem differences and item parameters

[†] Abe Yoshinobu, Takagi Masanori, Itchikawa Hisasi,
Faculty of software and Information Science, Iwate Prefectural University

3 関連研究

池田ら[5]は回答データの無い知識を問う多肢選択式問題において, 問題文の出題パターンと選択肢の類似度に着目した難易度推定方法を提案した. 本研究とは, 数学の確率論に関する問題を対象としている点で異なっており, 類似問題の計算過程に使われている数値の差に着目している点で異なる.

4 分析

本研究では, 2015年度から2019年度に本学で開講された情報基礎数学Bの授業内で実施された確認テストの問題を利用して分析を行った. 情報基礎数学Bの授業では最初に事前テスト, 最後に事後テストを行っており, このテストで出題された問題75問, 回答者数333名をもとにIRTの2パラメータロジスティックモデルを用いて項目特性を分析した.

4.1 基準問題と類似問題間での項目特性の相関

類似する問題どうしの識別力には正の相関があるという仮説を設定して分析を行った. 収集した75問の問題から問題文, 前提条件, 問い方, 問題文中の数値の4つの条件の中で一致する条件が1つ以上ある, かつ異なる条件が1つ以上ある問題を12組抽出し, 一方を基準問題, もう一方を類似問題としてこの2つの問題間の識別力の相関を求めたところ, 有意な強い正の相関 ($r=0.79$, $p=0.002$) があることがわかった. また, 困難度についても相関を調べたところ, 有意な相関は認められなかった. ($r=0.27$, $p=0.396$)

4.2 基準問題と類似問題間での困難度の差

4.1節では, 困難度で有意な相関が認められなかったため, 類似問題間で困難度にどのような差が出るか, また, どのような要素が困難度に影響を与えるか検証するため, 4.1節で分析対象とした類似問題12組で分析を行った. 分析結果を表1に示す. この中から困難度の差が大きい問題と小さい問題をそれぞれ抽出し要因を考察した.

4.2.1 困難度の差が大きい問題

問題4 (図2) では, 問い方が異なっており基準問題の方が発想の仕方が難しいと考えられるため困難度が高くなったと考えられる. また, 問題7 (図3) は複数の人を3人ずつ部屋に分ける場合の数を求める問題であったが, 全体の人数が多い問題ほど難易度が高くなることがわかった. また, 問題8も困難度の差が大きくなったが, 問題7と同様の理由

で困難度に差が生じていたため本稿では省略した。

表 1 基準問題と類似問題の困難度

問題番号	基準問題	類似問題	差
1	-1.846	-3.283	1.437
2	-3.681	-2.699	-0.982
3	-1.387	-1.66	0.273
4	-1.419	-2.753	1.334
5	-1.349	-1.549	0.200
6	-2.877	-3.16	0.283
7	-2.29	-5.35	3.060
8	-1.631	-3.638	2.007
9	-1.059	-3.052	1.993
10	-1.651	-1.559	-0.092
11	-2.173	-2.317	0.144
12	-1.545	-2.28	0.735

4.2.2 困難度の差が小さい問題

問題 5 (図 4) は、重複する文字を含むアルファベットの並びの場合の数を求める問題であり、どちらも数え上げで解を求めることができたため、困難度に差が生じなかったと考えられる。また、問題 10 (図 5) は、二項定理を用いて指定された x の累乗の項の係数を求める問題であり、回答の流れが固定されており、同じ考え方で解ける問題であったため困難度に差が生じなかったと考えられる。

基準問題

education の 9 文字を 1 列に並べるとき、母音と子音が交互に並び並べ方は何通りあるか。(2)

類似問題

education の 9 文字を 1 列に並べるとき、d と t の間に 2 文字入れる場合の数は何通りあるか。(2)

図 2 問題 4

基準問題

9 人の生徒を 3 人ずつ A, B, C の組に分ける方法は何通りあるか。

類似問題

6 人の生徒を 3 人ずつ A, B の組に分ける方法は何通りあるか。

図 3 問題 7

基準問題

a, a, b, c の 4 文字の中から 3 つ選んで 1 列に並べるとき、その並び方は何通りあるか。(2)

類似問題

a, a, b, b, c の 5 文字の中から、2 つ選んで 1 列に並べるとき、その並び方は何通りあるか。(2)

図 4 問題 5

基準問題

$(2x+1)^6$ の展開式における x^3 係数を求めよ。

類似問題

$(x+3)^6$ の展開式における x^2 の係数を求めよ。

図 5 問題 10

5 考察

各問題にはそれぞれ測りたい能力があるが (問題 4 では文章を読み取って並び替えの仕方を発想する能力など)、4.1 節の結果から、抽出した 12 組の基準問題と類似問題間でこの能力が同じであったため、測りたい能力が同じ問題では識別力が同程度になると考えられる。また、4.2 節の結果から、困難度に差が生じる条件として、答えを導く発想の仕方が異なることが挙げられる。この理由としては、基準問題と類似問題間で、発想に必要な知識や応用力に差が生じたためだと考えられる。また、困難度に差が生じない条件として同じ考え方で解けることが挙げられる。これは、同じ解法で解ける問題間には求められる能力に差が生じないと考えられるためである。

6 おわりに

本研究では、数学の確率論に関する類似問題と難易度の関係性を明らかにすることを目的とし、類似問題間の差異が項目特性に与える影響について考察した。その結果、類似問題間の識別力には強い正の相関があることや、答えを導くための発想が複雑になるほど困難度に差が生じ、数え上げができる問題や同じ考え方で解ける問題では困難度に差が生じないことが明らかになった。

しかし、今回の分析対象の問題には、授業の開始時点の事前テストと授業終了時点の事後テストの結果を含んでいるため、学習者が事前テストを解いたのちに当該問題を学習し、学習者の能力が向上したため困難度に差が生じたと考えられる問題もあった。そのため、今後は同時に問題を解かせたうえで困難度の差について分析していく。

参考文献

- [1] 中村洋一, "コンピュータ適応型テストの可能性", 日本語教育, VOL. 148, pp. 72-83 (2011)
- [2] 光永悠彦, テストは何を測るのか, ナカニシヤ出版, 2017
- [3] 加藤健太郎, 山田剛史, 川端一光, "R による項目反応理論", オーム社, 2014
- [4] Parshall, Cynthia G.; Kromrey, Jeffrey D.; Chason, Walter M.; Yi, Qing (1997) "Evaluation of Parameter Estimation under Modified IRT Models and Small Samples". Paper presented at the Annual Meeting of the Psychometric Society
- [5] 池田信一, 高木輝彦, 高木正則, 勅使河原可海, 多肢選択式項目の出題パターンと選択肢の類似性に着目した難易度推定方法の提案と評価, 情報処理学会論文誌, Vol. 54, No. 1, pp. 33-44, 2013