数理解析研究所講究録
第2022巻 2017年 143-148

項目応答理論を用いたオンライン問題の選別について

山口大学・教育学部　北本　卓也 (Takuya Kitamoto)
Faculty of Education,
Yamaguchi University

1　はじめに

小中学校、高等学校においてプログラミングの授業の導入が決められたことからわかるように、教育現場において ITC の導入が求められつつある。授業における ITC の活用はもとより、授業外においてもオンライン学習の導入が進みつつある地方公共団体もある。また、塾などの民間の教育機関の中にはオンライン学習を中心に教育を行っているところも存在する。このようなオンライン学習をうまく活用するための 1 つのキーは、オンライン学習に向いた良い問題を集めることである。そこで本稿では項目応答理論 (以下、IRT と書く) を用いてオンライン学習に向いている良い問題を選択することを検討する。具体的には、IRT で計算される各問題の識別力と難易度を用いて問題を選別する手法を考える。以下では、IRT と STACK、IRT を用いた問題選別の実験の順番に説明していく。

2　項目応答理論 (IRT) とは

項目応答理論 (以下、IRT と記す) は、TOEFL や IT パスポート試験などで実際に使われている能力判定方式である。通常の試験が正答加点方式であるのに対して、統計的な処理で受験者の能力を判定し、より精密な能力判定が可能であるとされている。

IRT では各問ごとに「難易度」と「識別力」というパラメータを設定する。難易度は問題の難しさを指している。識別力は受験者の能力を判定する力を示しており、識別力が高い問題というのは「能力が高い人はほとんど正解し、能力が低い人はほとんど不正解である」という問題であるである。いわゆる良い問題は識別力が高い問題という言い方ができる。IRT では、識別力が a_j、難易度が b_j である問題を能力値が θ である人が正解する確率 $p_j(\theta)$ は

$$p_j(\theta) = \frac{1}{1+e^{-Da_j(\theta-b_j)}} \quad (\text{ただし、}D = 1.7) \tag{1}$$

であると考える (これをロジスティックモデルという)。これより能力が θ である人が同じ問題を間違える確率は $(1-p_j(\theta))$ である。

IRT を用いるためには、各問ごとに設定される識別力 α と難易度 β を求める必要がある。この識別力と難易度を計算するためには、通常、「EM アルゴリズム」と呼ばれる特殊なアルゴリズムを用いる。EM アルゴリズムは生徒の能力値が正規分布に従うこと

を仮定し、収束計算で計算を行うがデータ量が十分多くないと収束しないことがある。今回の実験は1つのクラス(40名程度)で行ったため、収束計算が終了せず、通常の計算で初期値として用いる値で、この識別力と難易度の値を代用した。

識別力 α と難易度 β が求まれば、生徒の能力値 θ を次のようにして計算する。今、n 個の問題があり、j 番目の問題の (識別力, 難易度) が (a_j, b_j) で与えられるとする。この試験を受けた受験番号 i の人が j 番目の問題が正解した場合、$u_{ij}=1$ とし、不正解の場合 $u_{ij}=0$ と u_{ij} $(i=1,...,m,\ j=1,..,n)$ の値を定める。このとき、受験番号 i の人の能力値 θ の確率は

$$L_i(\theta)=\prod_{j=1}^{n} p_j(\theta)^{u_{ij}}\left(1-p_j(\theta)\right)^{1-u_{ij}} \tag{2}$$

で与えられる。これが現実に起こったということは、上の確率が最も高くなるように θ を設定するのが合理的ということになる。つまり、上の関数 $L_i(\theta)$ が最大値を取る θ の値を i 番目の受験者の能力値とする。

実際には上の $L_i(\theta)$ の対数をとった

$$\log\left(L_i(\theta)\right)=\sum_{j=1}^{n}\left\{u_{ij}\log\left(p_j(\theta)\right)+(1-u_{ij})\log\left(1-p_j(\theta)\right)\right\} \tag{3}$$

の関数の方が簡単なので (3) 関数 $\log\left(L_i(\theta)\right)$ を考える((2) の関数の最大値を取る θ と (3) の関数の最大値を取る θ は同じである)。関数の最大化を考えるには、(3) の関数の導関数 $\frac{d\log(L_i(\theta))}{d\theta}$ の零点を求めれば良い。

以上をまとめると、次の能力判定アルゴリズムを得る。

(i) 問題の識別力、難易度を求め、j 番目の問題の (識別力, 難易度) を (a_j, b_j) とする。

(ii) 試験結果を u_{ij} とし、(2) より θ の関数 $L_i(\theta)$ を定義する(ただし、受験番号 i の人が j 番目の問題に正解したときは $u_{ij}=1$、不正解の時は $u_{ij}=0$ とする)。

(iii) $\frac{d\log(L_i(\theta))}{d\theta}=0$ の θ に関する根を求め、それを i 番目の受験者の能力値とする。

3 STACKとは

STACK は Moodle のプラグインの1つとして使える E-learning システムである。STACK を用いると比較的容易に数学や理系の演習問題を作成することができる。最大の特徴は数式処理システム Maxima を内蔵していることにより数式を処理できることである。従来の E-Learning システムは数式を処理できないため、$x+1$ と $1+x$ を異なる答えとして認識してしまう。よって、E-Learning システムの解答方式として選択式しか使えない(現在使われているほとんどの E-Learning システムはそうである)。これに対し、STACK は数式処理システムを内蔵しており、$x+1$ と $1+x$ を同じ答えと認識する。このため、解答者に数式を入力させることが出来て、より柔軟な問題や解答方式が可能になる。例えば、乱数を用いて式の係数を替えることで、問題の形は同じであるが式が異なる複数の問題を1つの基本問題から作成することができる。

このように STACK を用いれば、多くの問題を自動的に生成することが比較的容易になるが、上で述べたように効果的な E-learning 学習のためにはオンライン学習向きの良い問題がキーとなるので IRT を用いた問題選択を考える。

4 IRT を用いた問題選別ついて

4.1 基本的な考え方

上で述べたように IRT における識別力が大きい問題とは、「能力値の高い生徒はほとんど正解し、能力値が低い生徒はほとんど不正解する問題」である。つまり、いわゆる良い問題とは「識別力が大きい問題」と考えられる。そこで次のようにして問題選別を行うことを考える。

(i) 問題を作成し、オンラインでテストを行い、正解・不正解のデータを得る。

(ii) 正解・不正解のデータをもとに IRT により、識別力・難易度を決める。

(iii) 識別力が小さな問題を排除する。

4.2 実験

微積分の授業の後、毎週、STACK を用いた E-learning の課題を課し、正解・不正解のデータを収集した。あくまで演習として E-learning の課題を出したので、何度もやり直せるようにした（乱数を用いて問題の数値を変えているので同じ問題は基本的には出ない）。このため、生徒の解答は複数あったが、最も得点が高かった時のデータを用いた。

このようにして収集したデータを用いて、識別力と難易度の計算を「EM アルゴリズム」で行ったがデータ量が十分でないため、計算が収束しなかった。そこで j 番目の問題の 識別力 α_j、難易度 β_j として下記の値を用いた。

$$\alpha_j = \frac{x_j}{\sqrt{1-x_j^2}}, \quad \beta_j = \frac{y_j}{x_j} \tag{4}$$

ここで x_j は受験者の正答率とその問題の正解・不正解ののデータとの相関係数、y_j はその問題の正答率である。下に実験結果を示す。

表 1: 問題セット 1

	問題 1	問題 2	問題 3	問題 4
α_j	0.242	1.06	1.07	1.62
β_j	−0.638	−1.33	−0.590	−0.0982

表 2: 問題セット 2

	問題 1	問題 2	問題 3	問題 4
α_j	0.513	2.24	1.18	1.29
β_j	-3.03	-0.943	-1.60	-0.852

表 3: 問題セット 3

	問題 1	問題 2	問題 3
α_j	1.92	0.893	1.92
β_j	-1.02	-2.33	-1.02

問題セット 1 は極限値を求める問題である。最も困難度の低い問題 1 は

$$\lim_{x\to 1}\frac{x^2-9x+8}{x^2-6x+5} \tag{5}$$

であるのに対し、最も困難度の高い問題 4 は

$$\lim_{x\to\infty}\left(\sqrt{x^2+7x}-\sqrt{x^2+5}\right) \tag{6}$$

であるが、両者とも解き方を知っていれば解ける問題であり、なぜこのような差が出てくるのか正直良くわからない。

問題セット 2 も極限値を求める問題である。最も困難度が低い問題 1 は

$$\lim_{x\to 0}\frac{f(a+3x)-f(a-3x)}{x} \tag{7}$$

であるのに対し、それと似ているが困難度がそれより高い問題 2 は

$$\lim_{x\to 0}\frac{xf(4a)-af(4x)}{3(x-a)} \tag{8}$$

である。こちらは問題 1 と比べると明らかに面倒な問題であり、この問題の困難度が高くなっているのはわかる。

問題セット 3 も極限も計算する問題であるが、こちらはロピタルの定理を活用する問題のセットである。問題 1 は

$$\lim_{x\to 0}\frac{1-\cos(2x)}{x\sin(5x)} \tag{9}$$

であるが、それより困難度の低い問題 2 は

$$\lim_{x\to +0}x^2\log(2x) \tag{10}$$

である。ロピタルの定理を活用する点から見ると問題 2 は分母と分子にわざと分けてからロピタルの定理を用いねばならない分難しく思えるが、実験の結果からは問題 2 の方

表 4: 問題セット 4

	問題 1	問題 2	問題 3	問題 4	問題 5	問題 6
α_j	0.465	1.24	0.509	0.983	1.51	1.33
β_j	−3.12	−1.30	−2.54	−1.88	−0.483	−1.65

が易しいということになっている。この問題は $x^2\log(2x)$ の x^2 のべき 2 と $\log(2x)$ の 2 を乱数で変えているが、乱数がどんな値になっても答えは 0 となってしまうため、何度か繰り返しやっていくうちに必ず答えが 0 になると生徒が覚えてしまったからではないかと思われる。

問題セット 4 は微分の問題である。難易度が低い問題 1，問題 3 はそれぞれ $f(x) = \sin^4 x$ と $f(x) = x^4\log(x)$ を微分する問題であり、最も難易度が高い問題 5 は $f(x) = \log|2-5x|$ を微分する問題である。これもそれほど難易度が違うようには思えないので、予測と実験結果が違っていることは予想外である。

4.3 考察

実験結果からは次のことが読み取れる。

- 学生にとっての問題の難易度は実験をやってみないとわからない。
- 難易度が低い問題（易しい問題）は識別力が高く、難易度が高い問題は識別力が高いという傾向がある。
- 乱数を用いて出題されるたびに異なる問題が出るようにしたが、乱数によって答えが変化しない問題は学生が答えを暗記するためか識別力が低くなってしまう。
- 今回は最高点の結果を解析に用いたが、問題選択の際には初回のテストの結果も考慮する必要がある。

5 結論

STACK を用いて、E-learning のオンライテストの演習を実施した。演習問題なので何度も繰り返しテストが受験できるようにしたが、乱数を用いて、毎回異なる問題が出題されるようにした。行ったオンラインテストの結果を用いて、各問題の識別力・難易度を求めた。

実際に出た結果は、予想されるものとかなり異なっていた。特に問題から困難度を予想することはかなり難しいということがわかった。難易度が低い問題は識別力が低く、難易度が高い問題は識別力が高いという傾向が見受けられたがあくまで傾向であり、絶対的なものではない。

識別力が低い問題はいわゆる良い問題であると思われるので、当初は識別力が低い問題を排除することを考えていたが、最高得点だけでなく、初回のテストの結果も考慮して問題選別を行う必要があることがわかった。また、今回の実験はデータの数が限られているので、データの信頼性が確かでない。実際の教育現場で同じように E-learning を行う場合もデータの数が限られてくることが考えられるので、そういった場合にデータの信頼性をチェックする仕組みが必要である。チェックする必要がある。

参考文献

[1] 豊田: 『項目反応理論 [入門編](第 2 版)』, 朝倉書店, 2012.

[2] 豊田 :『項目反応理論・理論編』, 朝倉書店, 2005.

[3] 豊田: 『項目反応理論 [中級編]』, 朝倉書店, 2013.

[4] 豊田: 『項目反応理論 事例編』, 朝倉書店, 2002.

[5] 大友 :『項目応答理論入門』, 大修館書店, 1996.

[6] 加藤, 山田, 川端 :『R による項目反応理論』, オーム社, 2014.

[7] 項目応答理論 (IRT) と EasyEstimation のページ: http://irtanalysis.main.jp/

[8] STACK 日本公式ホームページ : http://ja-stack.org/

[9] 中村, 中原, 秋山 :「STACK と Moodle で実践する数学 e ラーニング」, 数理解析研究所講究録, Vol. 1674, pp. 40-46, 2010.

[10] 中村 :『数学 e ラーニング』, 東京電機大学出版局, 2010.