

adaptive online IRT system について

桑 幡 隆 行^{†1} 作 村 建 紀^{†1} 廣 瀬 英 雄^{†1}

項目反応理論 (IRT) は、古典的テスト法と異なり、学生の能力だけでなく問題の難しさについての情報も得ることができる。adaptive online システムは、通常、事前に別のテスト結果から指定された item bank を用いることで学生の能力を評価するシステムである。不完全なデータであるため、同時に問題の難易度を求めることはできない。しかし、ここで新しく提案する Limit IRT (LIRT) を使用すれば、両方を同時に求めることが可能である。これは、予備の短時間のテスト、オンラインシステムなど、多くの状況に適用できる可能性がある。ここでは、新たに開発した adaptive online システムを、高校生のためのオープンキャンパスイベントに用い、わずか 5 問の数学問題を与えるだけでも学生の能力評価は正しく行なわれたことを紹介する。

adaptive online IRT system

TAKAYUKI KUWAHATA,^{†1} SAKUMURA TAKENORI^{†1}
and HIDEO HIROSE^{†1}

The item response theory (IRT) gives us the valuable information about the difficulties of problems as well as the abilities of students contrary to the classical test method. The adaptive online systems often use the problems and their difficulties via item banks provided in advance by different test results, and evaluate the students' abilities immediately after tests were performed. The systems usually do not give the problem difficulty values because of incomplete data. We have estimated the score in $[0, 1]$ to such unanswered problems by using a newly proposed limit IRT (LIRT), and given the students' abilities in addition to the problem difficulties recomputed. The newly developed online adaptive system worked with only five problems successfully. Therefore, this kind of system can be applied to many situations such as primary check systems, short-time tests, introductory online systems. We have used this system to an open-campus event to introduce the adaptive online IRT system to high school students. The application field was mathematics to the first grade in high schools.

1. はじめに

今日の一般的な試験、テストとして実施されるものの問題形式や配点は、製作者に委ねられていることが多く、数理的に問題の性質を見極める過程を経て、公正・正確さに重きを置くことは稀である。このような試験によって取得された学力の評価が一般的に利用されているが、人の一生に関わる重要な指標として扱われていることを考えると、看過してよい問題とは言えない。そこでこの問題を解消し、数理的な分析に基づいた公正な成績判断を目指して、試験理論として IRT (項目反応理論, item response theory; 以下 IRT とする) が提案されている¹⁾。

IRT は現代テスト理論とも称され、これ以前には古典的テスト理論というものが考えられてきた²⁾。古典的テスト理論は、項目の難しさや、項目が総点に与える影響の度合いなどを分析する理論であるが、この分析から得たパラメータは、受験者の特性の分布に完全に依存するため、それ以外の受験者集団への応用性が全くないという問題点がある。またそれ以前に、パラメータは出題するテストそのものに依存するため、テストの実施が行われた時点で、そのテスト自体の再利用性が大きく失われてしまうことも、この理論の問題のひとつである。

IRT では、受験者集団に依存しない特性を想定し、直接観測できない構成概念として捉え、潜在特性 θ を考える。この概念に基づいた項目パラメータを推定することによって、異なる受験者集団へ利用することで潜在特性を測定できる問題群を生成することができ、古典的テスト理論がはらむ問題点を解消することができる。

IRT は、TOEFL や J-CAT などの今日のいくつかの検定試験で実際に利用されている試験理論である。この理論は、テストを構成する項目 (問題) 一つひとつの特性を加味し、どの項目に正解したかにより評価を定めるもので、従来の素点による評価より正確な評価が得られる³⁾ ため、学生自身が習熟度を把握することにも適している。これによって問題の本質的難度を明確にすることが出来れば、被験者の能力に応じた問題を適宜提供するといった適応性のあるシステムの構築も可能であると考えられる。また、適応型試験で得られる反応パターンは、不完全な欠損マトリクスであるため、項目パラメータ推定には使えない。そこで、IRT を用いて欠損を補填し完全マトリクスを生成する手法を提案することで、適応型試験

^{†1} 九州工業大学
Kyushu Institute of Technology

結果を用いた試験システムを提案する．

しかし，IRT を利用した学習システムは，日本において普及していない．これは伝統的な素点評価の意識が根強いことや，そもそもの IRT の知名度の問題も原因に挙げられる．学校教育機関に限らず，会社内における教育支援といった用途など，IRT を導入できる環境を模索することも課題のひとつである．IRT を組み込み，動的に短時間で高精度の受験者特性の測定を行うことを目的としたコンピュータ適応型テストを特に CAT (Computer adaptive-testing) と呼ぶ．

本研究では，アダプティブオンラインテストの実現を目指し，試験的なシステムを構築，実施して効果を確認することを目的とする．利用した理論である IRT と，実際に構築したシステムおよびそのインタフェースを解説し，最後にテストを実地運用して得られた知見と有効性を述べる．

2. adaptive テストに向けた項目パラメータの推定過程 - IRT (項目反応理論)

本研究で扱う IRT に関して，どのように利用するか含め，解説する．各パラメータの推定方法には様々なものが使われており，一般的に広く使われているのは，最尤推定法を用いることである．

2.1 古典的テスト理論

$j = 1, 2, \dots, n$ 問の問題に，学生 $i = 1, 2, \dots, N$ が回答したとする．回答が正答なら 1，誤答なら 0 を与える．これによって得られた図 1 のような $N \times n$ のマトリクスは，反応パターンと呼ばれる．

古典的テスト理論における項目分析の指標の一つが通過率である．通過率は正答率とも呼ばれる． N 人の受験者の中で項目 j に正答した受験者が k_j 人いたとすると，通過率は

$$p_j = \frac{k_j}{N} \quad (1)$$

と定義される．反応パターン $\sigma_{i,j}$ は 0/1 データであるため，

$$p_j = \frac{1}{N} \sum_{i=1}^N \sigma_{i,j} \quad (2)$$

とも表現することができる．

通過率の次に計算される項目分析の指標は識別力である．識別力が高い項目は，項目得点がテスト全体で測っている特性を適切に反映し，受験者を区別している項目である．逆に識別力の低い項目は，項目得点が特性を適切に反映しない項目である．全 n 問の問題群にお

	item																			
user01	1	1	1	0	1	1	1	1	0	0	1	1	0	1	1	1	1	1	1	0
user02	1	1	0	1	1	1	1	0	0	1	1	0	0	1	0	1	1	1	1	0
user03	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
user04	1	1	0	1	1	1	1	1	1	1	0	1	0	1	0	1	1	1	1	0
user05	1	0	0	0	0	1	1	1	0	1	0	1	0	0	0	1	1	0	0	0
user06	1	1	0	1	1	1	1	0	1	0	0	1	0	0	0	0	1	0	0	0
user07	1	0	1	1	1	1	1	0	1	1	0	0	0	1	1	1	1	0	1	0
user08	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1
user09	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	1	1	0
user10	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1
user11	0	0	0	0	1	1	1	0	1	0	1	1	0	0	0	1	1	1	1	1
user12	1	1	0	1	1	1	1	1	0	1	0	1	1	0	1	1	1	0	1	1
user13	1	1	1	1	1	1	1	0	1	1	1	0	0	1	1	1	1	1	0	1
user14	1	1	0	1	0	0	0	1	0	1	1	0	1	0	0	1	1	0	1	1
user15	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	0	1	0

図 1 反応パターン

Fig. 1 Reaction pattern

ける j という問題に w_j という配点が与えられているとき，受験者 i のテストの得点は

$$y_i = \sum_{j=1}^n w_j \sigma_{i,j} \quad (3)$$

で計算する．

項目 j の識別力 r_j は， σ_j と y の，反応パターンとテスト得点の相関係数

$$r_j = \rho(\sigma_j, y) \quad (4)$$

で定義される．

識別力は，最大値が 1 であり，1 に近いほど識別力が高く，0 に近いほど識別力が低いと判断する．

2.2 IRT (項目反応理論)

IRT では，被験者ごとに直接観測できない潜在特性値である能力値 θ を横軸にとった ICC (項目特性曲線, item characteristic curve) によって，正答確率 $P(\theta)$ を表すことができる．今回は，問題 j と受験者 i に関して，能力値パラメータ θ_i と，2 つの項目パラメータである困難度 b_j ，識別力 a_j が関数となる 2 パラメータのロジスティックモデルに，正答確率 $P_j(\theta_i)$ が従うと考える． θ_i は学生 i の能力パラメータと呼ばれており，大きいほど能力が高いことを示す．困難度 b_j がより大きい場合，その項目はいわゆる難しい問題であり，全ての能力において正答確率が低くなる．識別力 a_j がより大きい場合，その項目は能力に応じた正答率の変化が大きくなる．

2.3 パラメータ推定

2 母数ロジスティックモデルでは、学生 i が問題 j に正解する確率 $P_j(\theta_i)$ とし、問題特性を以下のように表すことができる³⁾。

$$P_j(\theta) = \frac{1}{1 + \exp(-1.7a_j(\theta_i - b_j))} \quad (5)$$

試験結果に対する実現確率（尤度）は (6) で与えられる。

$$L = \prod_{i=1}^N \prod_{j=1}^n P_j(\theta_i)^{\delta_{i,j}} \cdot (1 - P_j(\theta_i))^{1-\delta_{i,j}} \quad (6)$$

ここで、 θ_i , b_j , a_j はすべて未知である。(6) の尤度 L を最大にする能力パラメータ θ_i と項目パラメータ b_j , a_j を求める。この方法を同時最尤推定法という。

尤度関数から能力値パラメータを積分消去して得られる (7) 式より項目パラメータ b_j , a_j を求める。これを周辺最尤推定法³⁾⁵⁾ という。

$$L(a_j, b_j) = \prod_{i=1}^N \left[\int_{-\infty}^{\infty} g(\theta) \prod_{j=1}^n P_j(\theta)^{\delta_{i,j}} \cdot (1 - P_j(\theta))^{1-\delta_{i,j}} d\theta \right] \quad (7)$$

そして、周辺最尤推定法で得られた項目パラメータをもとに、受験者の能力値パラメータ θ_i を推定する。アダプティブテストは、能力に応じて出題するという面から、得られた項目パラメータ b_j , a_j によって出題する問題を選定しなくてはならない。今回は、得られた反応パターンに応じて適宜能力値 θ_i を算出し、それに最も近い困難度 b_j をもつ問題を出題することで、適応性を実現する。

3. アダプティブオンラインテスト

これまでの研究では、LMS (Learning Management System) である Moodle への IRT 解析システムの組み込みを試みてきたが、本研究では、Moodle を使用せず、プログラミング言語 PHP (PHP: Hypertext Preprocessor) による構築を目指した。

3.1 IRT 解析システムを組み込んだ Moodle テストシステム構築 - これまでの研究

Moodle は、PHP によって構築されたオープンソースであり、ユーザ、コース、活動といった概念で構成されている。PHP の知識があれば、プラグイン作成やシステム改良が可能といった特徴や、モジュールやプラグインを導入することによって拡張できる。また、標

準搭載されている小テストモジュールを利用すれば、多肢選択問題や記述問題など、さまざまなタイプの小テストを設計し設置することができるため、これを利用することでテストシステムが実現できる。

しかし、Moodle にはユーザズマニュアルがなく、また拡張性を維持するために、コードがあらゆる方面に渡って編集可能である。更に、日本語システムの構築を目指す場合、英語システムの変数名を日本語モジュールに対応させたリンクを設計しなければならないといった不都合な点も見受けられた。これまでの研究によって、Moodle 改良によるシステム構築には、高い PHP 技術と大きな労力を要することがわかっている⁴⁾。これを受けて本研究では、PHP を用いて抜本的にシステムを構築するものとする。

3.2 テストの概要

本研究のテストシステムは、高い汎用性と伝達性を実現するために、Web ブラウザ環境だけで動作するよう、統合オープンソース xampp (ザンプ) を利用してサーバを立ち上げ、PHP ページで製作する。ただし、実際に能力値を算出するのは、特定の PHP ページが呼び出す Java プログラムが行う。

PHP は、動的に HTML ページを生成して Web ページを実現するプログラミング言語の 1 つである。動作は Web 上の応答に限らず、場合によってはサーバ側のファイルを読み込んで実行する他、ファイル出力、データベース操作など多岐に渡り、汎用性が高い。問題点としては、Web 用のソースであるためコンパイルが必要なく、完全にコードを秘匿して配布することが難しいことや、その脆弱性などが挙げられるが⁶⁾、本研究においては、配布や外部からの攻撃といったことは特に問題としない。今回は Web 上での動作性と、汎用性の高さから採用した。外部 Java に算出を委ねる理由は、処理速度が PHP よりも高速な上、複雑な推定を行うプログラムコードが Java で構築済であるからである。また、膨大なユーザの微細なデータを処理・保持するため、xampp に同梱されているデータベースアプリケーション MySQL も併用する。

3.3 テストの特徴

本システムの特徴として、Web ブラウザさえあれば動作するため、新規にアプリケーションを導入する手順が必要無く、マシンや OS にほとんど依存せずにテスト受験を開始できる点が挙げられる。また、動的に選ばれたごく少数の問題に回答するだけで、高精度の能力評価が行えるという CAT 特有の利点も備えている。

また、新規に作成した項目をすぐに運用に導入することはできない。新しい問題群を利用する場合には、いくつかの受験者 n から得た全問題 N への回答から、 $N \times n$ の反応パタン

を作成し、予め問題パラメータを推定しておくことで、初めて導入可能となる。推定は、前項に述べた最尤推定法に基づいた手法で行う。その回答ボタンを取得するために実施するテストが、次項にて述べるプレテスト（模擬試験）である。

3.4 プレテスト

プレテストは、このような CAT において、困難度 b と識別力 a を入手するために必要となるテストである。すべての項目に対する回答から得た反応パターンから、各受験者の能力値 θ と、困難度 b 、識別力 a を推定する。

今回のプレテストは、九工大生 15 人に対して実施した。使用した問題は、高校生向けの参考書から作成した、一般高校数学問題 30 問で、内訳は「数と式・二次方程式・三角関数」11 問、「角度・三角比・図形」8 問、「確率統計」11 問となっている。以下に取得された問題がもつパラメータをジャンル毎に記載する。

表 1 各ジャンルデータベース
Table 1 Database of each genre

q_math(数と式・二次方程式・三角関数)				q_shape(角度・三角比・図形)				q_statistics(確率統計)			
id	ans	b	a	id	ans	b	a	id	ans	b	a
1	4	-1.489	0.958	1	4	-0.423	0.649	1	4	-1.017	1.007
2	3	-0.566	0.826	2	1	-0.233	1.077	2	4	-1.011	0.672
3	3	0.157	1.444	3	2	-0.987	1.049	3	2	-0.205	1.194
4	3	-0.824	0.853	4	2	-1.176	0.836	4	2	1.049	0.811
5	4	-0.801	1.409	5	2	0.339	0.866	5	2	-0.987	1.049
6	1	-0.7	1.035	6	2	-0.624	0.747	6	3	-0.056	0.982
7	2	-1.29	1.182	7	4	-0.352	1.387	7	1	-0.143	1.612
8	1	-0.604	1.247	8	2	-0.679	1.074	8	1	0.002	1.464
9	4	0.657	1.293					9	2	0.355	0.693
10	1	-0.364	0.739					10	2	-0.454	1.045
11	1	-0.2	1.217					11	2	1.259	0.575

4. アダプティブオンラインテストの構成

主なコーディングには、フリーソフトウェアの EZhtml を活用した。PHP のバージョンは最新の 5.3 である。

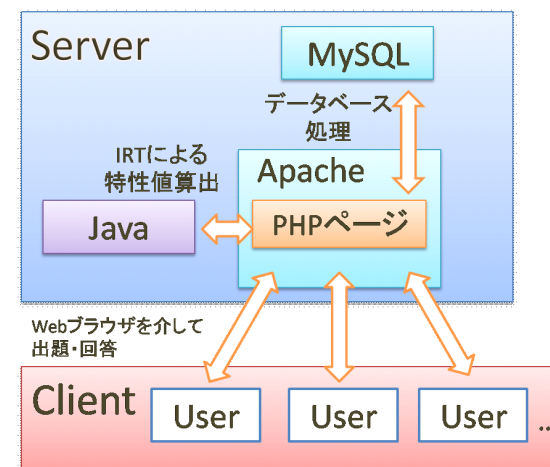


図 2 システム概要図

Fig. 2 System conceptual diagram

4.1 テストの実装仕様

図 2 に、本システムの概要図を示す。まず、xampp に同梱されているサーバアプリケーション Apache が IP アドレスによる Web スペースを提供し、PHP ページを公開する。受験者は、各マシンの Web ブラウザからページにアクセスし、参加名を半角英数字で入力することで試験を開始することができる。このとき、参加人数が最大 30 人程度であれば、サーバマシンは一勢アクセスに耐え得る。受験中は、選択したジャンル、回答番号に基づき、動的にページを変化させることで Web テストが進行していく。同時にバックグラウンドでは、被験者の回答データのデータベース収納や、回答ごとに適宜行う能力値算出等の処理を行なっていく。能力値は、前述の IRT 理論に基づいて導出され、その能力値に適応した問題が順次出力されていく。途中で中断することも可能であり、その場合は回答が完了している問題の結果が出力される。

PHP ページ、MySQL、IRT 解析プログラムの連携の流れは以下のようになる。

- (1) 識別 ID を受験者に入力させ、受験者毎に MySQL でテーブルを作成する。
- (2) 受験者の選択によって、受験者に出題・照合する問題テーブルを決定。
- (3) 第一問を画像で出題。

- (a) 回答番号を問題テーブルと照合して正否判断し、受験者テーブルにデータを追加
 - (b) IRT 解析プログラムにパターンを入力して実行し、その時点での能力値を示す
 - (c) テーブルの未出題問題群の困難度と能力値を比較、最も差の絶対値の低い問題を出题
- (4) (a)~(c) を指定回数だけ繰り返す (5 回) か、途中終了ボタンを押すと、結果画面に移行
- (5) 生成された受験者テーブルの各要素に基づいて、結果画面を生成し出力

4.2 インタフェースの流れ

実際に運用する場合には、コンピュータテスト全般の問題として、微妙な正否判断や、その部分点に関する問題を考慮しなければならない。例えば「～について簡潔に説明せよ」といった記述式の問題の採点は、採点者の裁量に左右される部分が否定できない。このため、一般にコンピュータテストの出題形式は、簡単のため選択形式が採用される。

今回作成したテストは、高校生向けの数学テストである。試験開始時に、受験したい問題ジャンルをクライアントである受験者が選択し、試験を開始する。回答は 4 択の選択形式で行い、試験が終了すると結果が表示され、 θ の変化の様子と、その受験者が解いた問題の解答と解き方を表示する。インタフェースは、大きく分けて ID 入力画面、ジャンル選択画面、問題回答画面、結果画面の 4 つのページで構成される。

4.2.1 ID 入力画面

図 3 に、ID 入力画面を示す。ユーザ ID を入力し、参加登録を行う。ここで入力されるユーザ ID がデータベーステーブル名となり、ジャンル選択画面に移行する際に新規作成される。実際には、テーブルの重複作成や参照の問題を避けるため、入力した ID の末尾に登録時の日本時間を追加する。

4.2.2 ジャンル選択画面

図 4 に、ジャンル選択画面を示す。登録が完了すると、この画面に移行する。受験に関する注意が示しており、問題ジャンルを選んで試験を開始する。問題ジャンルは 3 つに分かれている。PHP 処理によって、ここで選ばれた問題テーブルの使用を決定する。

4.2.3 問題回答画面

図 5 に、問題回答画面を示す。問題ジャンルを選択すると、この画面に移行する。問題の出題、回答を行う、このテストシステムの実質的な核となる画面である。

1 問目は、各ジャンルの問題を困難度順にソートし、中心の困難度 b を持つ問題を選び出



図 3 ID 入力画面
Fig.3 ID input screen

題する。総問題数が偶数個の場合は、中心より困難度が 1 つ低いものが選択される。1 問回答すごとに、IRT により能力値 θ を算出し、それに最も近い困難度 b を持つ問題が出題され、以降それを繰り返す。2 問目以降は、前問の回答の正否と、困難度が上昇したか否かを表すアイコンが表示される。能力値の精度が一定値に達するまで出題を繰り返すのが理想的であるが、今回は試験的システムであるので、最大でも 5 問回答した時点で終了する。また、下の終了ボタンから任意のタイミングで試験を終了し、それまでの解答のみで結果画面へ移行することも可能である。回答済みの問題はアイテムバンクから除外され、出題されない。

4.2.4 結果画面

図 6 に、結果画面を示す。5 問解答するか、任意終了するとこの画面に移行する。

最終的な能力値をしきい値判断することによってランクを設けているため、これによるランク評価と、回答に応じて順次測定した能力値 θ がどのように変化したかを表すグラフを表示する。ランク評価の分類は、ランク S: $0.5 \leq \theta_i$, ランク A: $0.0 \leq \theta_i < 0.5$, ランク B: $-0.5 \leq \theta_i < 0.0$, ランク C: $\theta_i < -0.5$ となっている。また、問題の正誤と解法も同時に表示する。

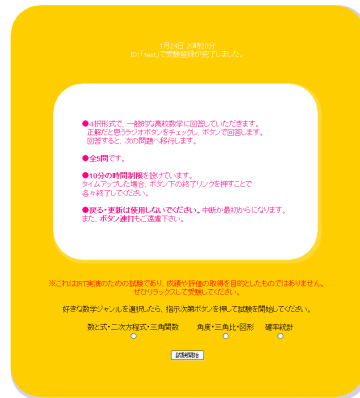


図 4 ジャンル選択画面
Fig. 4 Genre selection screen

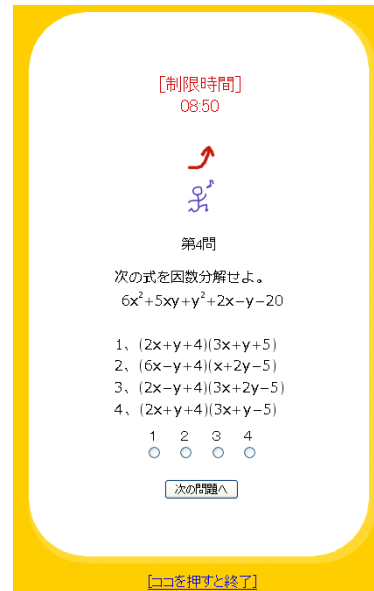


図 5 問題回答画面
Fig. 5 Answer screen

5. アダプティブオンラインテストの実地運用

今回は、オープンキャンパスに来場した高校生を対象に実施した。前述したとおり、搭載した問題はプレテスト実施済みの高校数学問題 30 問である。

5.1 運用結果

図 7～10 に、被験者の能力値 θ_i と困難度 b_j の変化の様子を示す。横軸が問題番号を示し、マーク付きが回答の正誤を表す困難度 b_j のグラフ、マークなしが能力値 θ_i のグラフである。

能力値 θ_i と困難度 b_j は、基本的に同じ変化がみられるが、これは能力値に応じて問題を出題するという当テストの仕様から自然に導かれる反応である。しかし一部には、正解後に困難度が低下する、または不正解後に困難度が上昇するといった反応を示し、能力値 θ_i と困難度 b_j の変化にずれが見られるサンプルもある。これは、連続した正答若しくは誤答によって、適当な問題がアイテムバンクから見つからなかったことが原因と考えられる。この

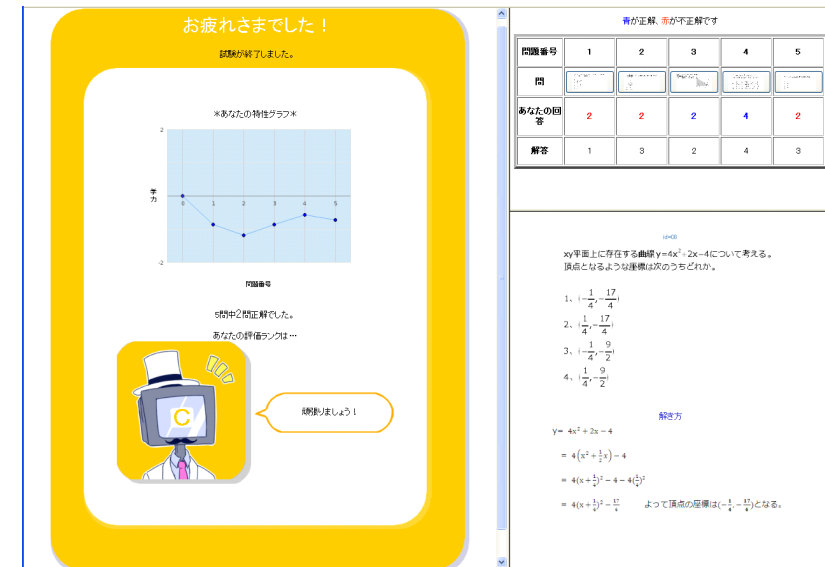


図 6 結果画面
Fig. 6 Result screen

ような反応は、同様の理由から、試験の進行に従って発生する可能性が上昇していくと考えられる。

6. 成長型システムのための不完全マトリクスの補完手法の提案 - LIRT

アダプティブテストを運用して得られるデータは、受験者の能力値 θ を推定することに役立つ。しかし、この測定データは、プレテストで取得した反応バタンのような完全マトリクスではないため、これだけのデータから項目パラメータ b, a を推定することはできない。だが、アダプティブテストで得られたデータを問題群へ還元することができれば、項目パラメータが逐次更新される成長型システムが実現できると考えられる。本項では、アダプティブテストより生成した不完全マトリクスを補完する LIRT (Limit IRT) という手法を提案する。

6.1 LIRT

用意された不完全マトリクスは、前項の実地試験の受験者 176 人から得たものである。こ

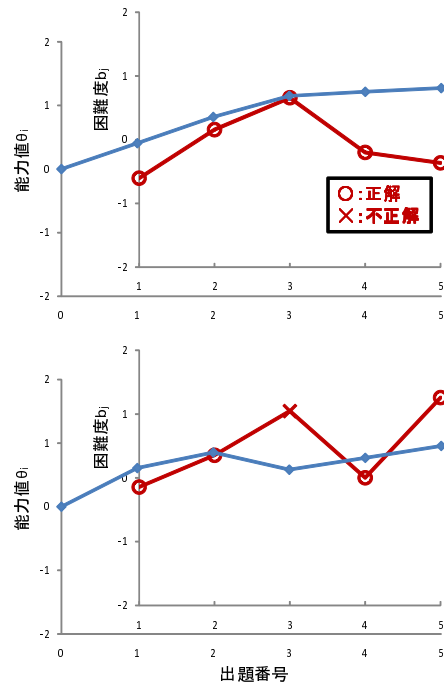


図 7 θ_i と b_j の変化の様子 (S ランク)

Fig. 7 state of the change of θ_i and b_j (rank S)

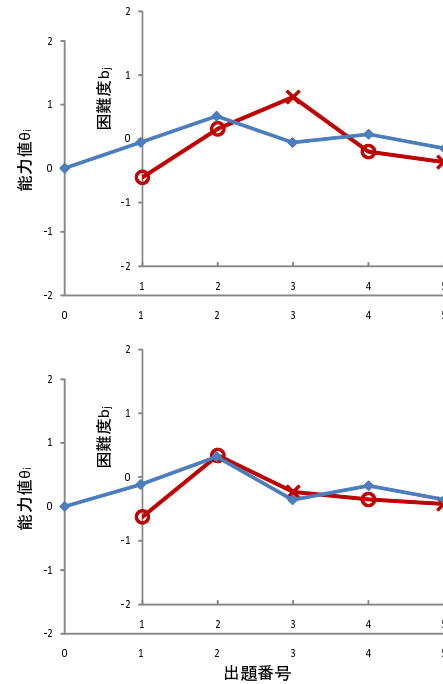


図 8 θ_i と b_j の変化の様子 (A ランク)

Fig. 8 state of the change of θ_i and b_j (rank A)

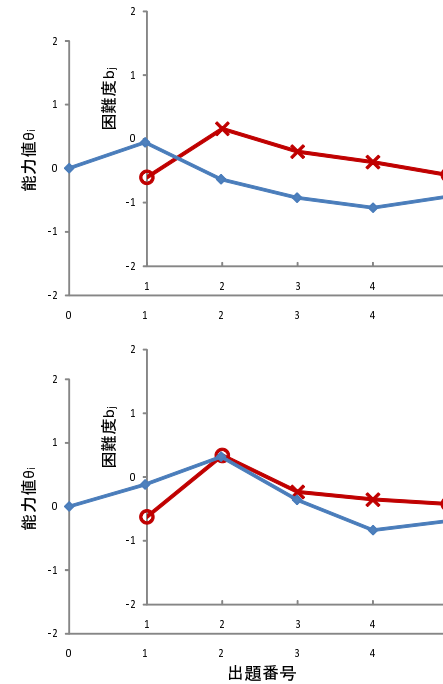


図 9 θ_i と b_j の変化の様子 (B ランク)

Fig. 9 state of the change of θ_i and b_j (rank B)

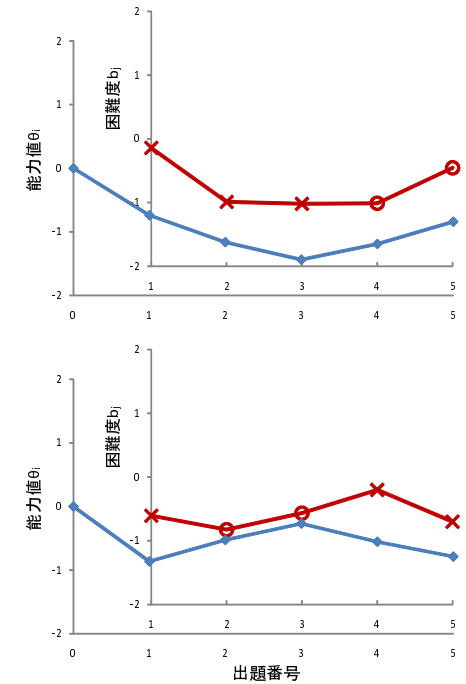


図 10 θ_i と b_j の変化の様子 (C ランク)

Fig. 10 state of the change of θ_i and b_j (rank C)

のマトリクスに対し、まず各受験者の平均点で空欄を埋めたマトリクスを生成し、元のマトリクスの非欠損パターンを上書きすることで、即時的な完全マトリクスを生成する。更にこのマトリクスに対し IRT に基づくパラメータ推定を行い、 θ_i, b_j, a_j を算出し、2 母数ロジスティック式 (5) によって正答確率 $P_i(\theta_j)$ を求め、新たな完全マトリクスを生成する。これ以降は、元マトリクスからの上書き、パラメータ推定を、マトリクス全体が収束するまで (限界まで=Limit) 繰り返す。元の不完全マトリクスから、図 11 のような完全マトリクスが得られた。このようにして完全マトリクスを生成する手法を LIRT として提案する。

7. ま と め

最大でも 5 問というごく小規模なテストながらも、適応性のあるテストが実施され、能力測定が実施することができたといえる。また、出題中表示されるアイコンから、受験者側で能力評価を確認しながら学習を行うことができたため、今回の目標を達成することができた。

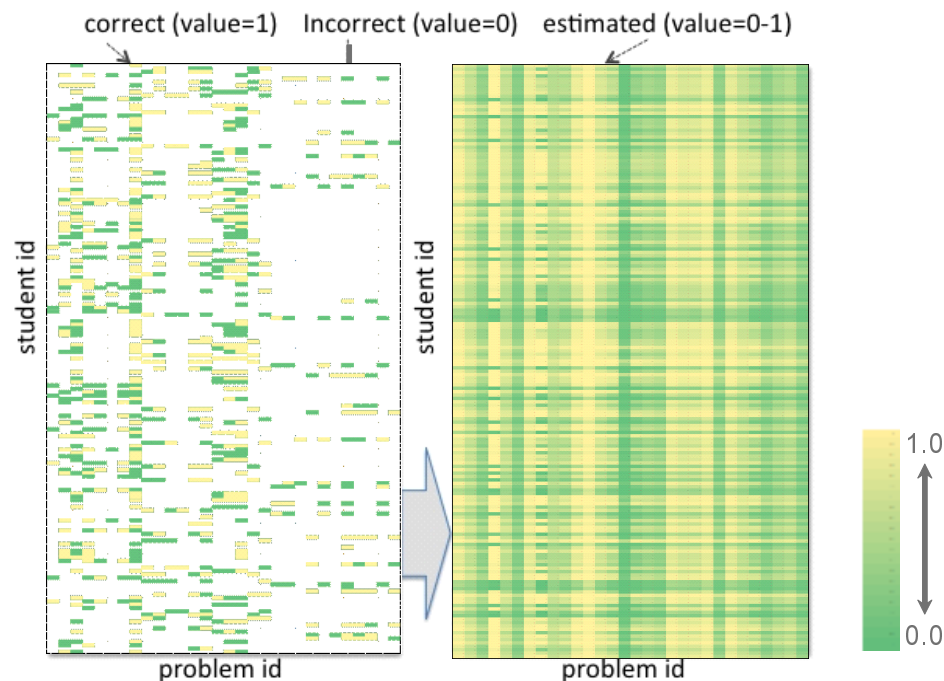


図 11 LIRT によるマトリクス補完
Fig. 11 Matrix completion by LIRT

謝辞 本論文を執筆するにあたり、PHP コーディングに関して多くの助言を下された中園暢さん、並びに当研究室の方々に、厚く御礼申し上げます。

参 考 文 献

- 1) Frank B. Baker, Item Response Theory: Parameter Estimation Techniques, Marcel Dekker, Inc(1992)
- 2) 池田 央, 現代テスト理論 (1994)
- 3) 豊田 秀樹, 項目反応理論 [入門編] テストと測定の科学 (2002)
- 4) 山口 裕佑, IRT システムの Moodle への実装 (2010)
- 5) 作村 建紀, 項目反応理論を拡張した最適能力判定システム (2010)
- 6) PHP: PHP マニュアル - Manual <http://php.net/manual/ja/index.php>
- 7) H. Hirose, Validities of the Signed and Unsigned Lecture Questionnaires Using the Item Response Theory, US-China Education Review , Vol.1, No.5, pp.627-635 (2011)
- 8) T. Sakumura, H. Hirose: Test evaluation system via the Web using the item response theory, Information, Vol.13, No.3, pp.647-656 (2010)
- 9) H. Hirose: An optimal test design to evaluate the ability of an examinee by using the stress-strength model, Journal of Statistical Computation and Simulation, Vol.81, No.1, pp. 79-87 (2011)
- 10) 月原, 鈴木, 廣瀬: 項目反応理論による評価を加味した数学テストと e-learning システムへの実装の試み, コンピュータ&エデュケーション (CIEC) , Vol.24, pp.70-76 (2008)
- 11) T. Sakumura, T. Kuwahata and H. Hirose, An Adaptive Online Ability Evaluation System Using the Item Response Theory, Education& e-Learning (EeL2011), pp.51-54, November 7-8, (2011)
- 12) T. Sakumura and H. Hirose, Student Ability Evaluation using the Stress-Strength Model When Ability is the Random Variable, Proceedings of the 2010 International Congress on Computer Applications and Computational Science (CACCS 2010), pp.865-868, December 4-6, (2010)
- 13) H. Hirose and T. Sakumura, An Accurate Ability Evaluation Method for Every Student with Small Problem Items Using The Item Response Theory, Proceedings of the International Conference on Computer and Advanced TEchnology in Education (CATE 2010), pp.152-158 (2010)
- 14) T. Sakumura and H. Hirose : Test Evaluation System via Web using the Item Response Theory, "The Fifth International Conference on Information" INFO2009, pp.112-115, November 6-9, (2009)