

項目応答理論を使った学生の能力推定と それに対応した教材選択手法の開発

福井県立大学 学術教養センター
山川 修
osamu@laputa.fpu.ac.jp

[アブストラクト]

福井県立大学では、項目応答理論(IRT)を使い、新入生のコンピュータ利用能力(スコア)を継続的に測定しているが、IRTをベースに推定されたスコアに対応する教材を選択する方法を考案したので、その解説を行う。通常、能力に応じた教材を提供するためには、インストラクショナルデザイン(ID)に基づいて教材をトップダウンにデザインする必要があるが、ここで考案した方法は、教材を学習した学生にアンケートを実施し、その結果と学生のスコアをもとに、各教材の「知識度」を推定するというボトムアップで簡便な方法である。

[キーワード]

項目応答理論, IRT, 能力推定, 教材選択手法, 知識度

1. はじめに

高等学校において教科「情報」が必修になり、コンピュータ利用の基礎的知識を、学生は大学入学時までに獲得してくることが期待されている。そのため、大学におけるコンピュータ利用教育(以下、「利用教育」と略す)の見直しが必要であるという議論が多くのごとくでなされている[1]。学生は大学在学中の勉学の中で、レポートをワープロで作成したり、課題を電子メールで提出するなど、基本的なコンピュータ利用能力(以下、「利用能力」と略す)が必要とされる。そのため、最低限の利用能力を獲得していない学生は、勉学に支障をきたすことになる。また、大学の新生に、毎年利用教育を実施している感触では、全体的に学生の利用能力は向上しているが、その広がりは大きくなっているようである。そこで本学では、従来から実施されている利用教育を一律に廃止するのではなく、学生の入学時点での利用能力に応じたリメディアル教育を実施し、大学で必要とされる利用能力を全学生が獲得できる方向で、利用教育の見直しを検討している[2]。

本プロジェクトでは、上記の見直しを実現するために必要となる、新入生の利用能力を安定的に測定する方法と、測定された能力に対応する教材を提供する方法について研究を行った。その結果、利用能力に関する簡単な質問から項目応答理論(Item Response Theory: IRT)を使って各学生のスコア(能力値)を推定した結果、安定的に学生の利用能力の測定が可能ながわかった。また、学生の利用能力に対応する教材に関しても、簡単なアンケートから学生の能力値に対応した教材の「知識度」を定義できることがわかった。

2. 項目応答理論(IRT)

項目応答理論(IRT)は、項目反応理論とも呼ばれるが、テストの難易度に依存しない尺度で被験者の能力を推定したり、テスト項目の特徴を捉えたりするための統計理論である[3][4]。IRTでは、等化

という手順を踏むことにより、一部のテスト項目を変更して別の集団にテストを実施した場合でも、共通の尺度上で被験者の能力を推定することが可能になる。そのため、今回のように、継続的に大学入学生の利用能力を測定する場合、適していると考えられる。

IRTには、テスト項目に対する被験者の能力値（特性値）と正答確率の関係を表現するために、様々なモデル（項目特性関数）が存在するが、本研究では、2母数ロジスティックモデルを採用した。2母数ロジスティックモデルにおける項目特性関数は以下のように表現できる。

$$P_j(\theta) = \frac{1}{1 + \exp(-Da_j(\theta - b_j))}, \quad -\infty < \theta < \infty$$

ここで、 $P_j(\theta)$ は項目 j に正答する確率、 θ は被験者特性値（能力）、 $\exp(\)$ は指数関数、 D は定数1.7で、正規累積モデル[5][6]との対応を取るための数値である。また a_j と b_j は項目 j を特徴づける母数で、 a_j は識別力を示し、 b_j は困難度を表す。

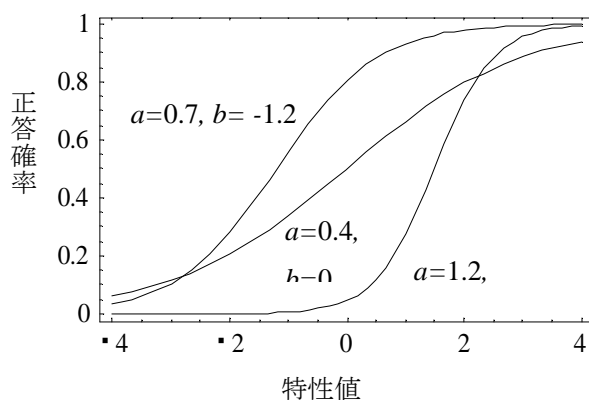


図1 識別力と困難度が違う場合の項目特性曲

識別力 a と困難度 b が違う3種類の項目特性曲線の例を図1に示す。困難度が大きい場合、曲線は右にずれ、識別力が大きい場合、曲線の中央付近の傾きが急になることがわかる。

3. 能力値の測定

本学における利用教育の詳細は、参考文献[2]を参照されたいが、2004年度～2007年度の3年間、利用教育の最初（プレースメントテスト：以下、「pテスト」と略す）と最後（アチーブメントテスト：以下、「aテスト」と略す）で、日本語入力、Windows操作、インターネット利用の、3分野の項目からなるテストを実施した。項目数は各分野10問で合計30問、すべて5つの選択肢からの択一式問題である。なお、pテストとaテストでは同じ問題を出題した。ただし、2005年度に実施する際に、2004年度の問題項目の20%程度を変更した。これは、極端にやさしい項目、および、識別力が低い項目を除き、新しい項目を加えたためである。

表1にIRTの解析ソフトであるBILOG-MG3を使って推定した各項目の困難度と識別力を示す。ただし、母数に適当な線形変換を施してあるので、図1の数値とは大きくずれている。2004年と2005年度のテストに採用されていた項目は○印で示してある。×印はその年のテストでは採用されなかったテスト項目である。なお、2006、2007年度は、2005年度と同じテスト項目を使用した。テストを受験したのは、2004年度pテスト362名、aテスト224名、2005年度pテスト375名、aテスト365名、2006年度pテスト382名、aテスト339名、2007年度pテスト391名、aテスト361名であった。

図2に、テスト項目の困難度と識別力を横軸と縦軸にして、各項目の値をプロットしたものを示す。

図中、丸印で囲まれている点は、2005年度のテストの際、除かれた項目である。識別力が低いか、困難度が低い項目が除かれていることがわかる。

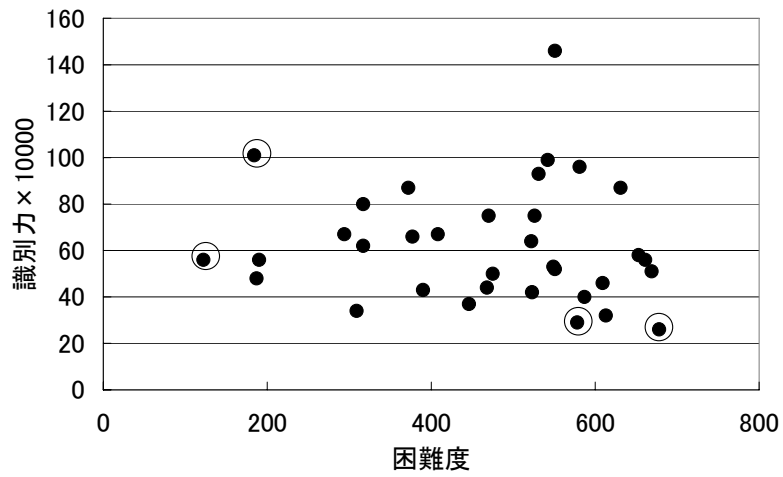


図2 各項目の困難度と識別力

表1 テスト項目の変更

	項目名	困難度	識別力 ×10000	2004	2005
日本語入力	確定キイ	190	56	○	○
	大文字への一時切替	187	48	○	○
	変換キイ	317	62	○	○
	BS キイと DEL キイ	390	43	○	○
	変換中の訂正	446	37	○	○
	変換後の訂正	587	40	○	○
	Num-Lock	549	53	○	○
	カタカナ入力	551	52	○	○
	Cap-Lock	609	46	○	○
	日本語入力の切替	468	44	○	○
Windows 操作	右クリック	317	80	○	○
	スタートボタン	294	67	○	○
	ドラッグ	372	87	○	○
	ファイルのプロパティ	653	58	×	○
	Window の位置変更	470	75	×	○
	ファイルの種類	542	99	×	○
	コントロールパネル	661	56	×	○
	ファイルとフォルダ	531	93	○	○
	Window のサイズ変更	377	66	○	○
	ファイル名の拡張子	551	146	○	○
	クリック	122	56	○	×
	ハードディスク	578	29	○	×
	ドライブ	678	26	○	×
	インターネット利用	文字化け	408	67	○
添付ファイルの送付		475	50	○	○
Web ページ		523	42	○	○
CC の役割		526	75	○	○
メーリングリスト		522	64	○	○
添付ファイルの受信		613	32	○	○
クッキー		631	87	○	○
HTML		581	96	×	○
電子メールアドレス		309	34	○	○
ネットワーク設定		669	51	×	○
ダウンロード	184	101	○	○	

図3に、問題項目の例（Caps Lock）を示す。図に示すように選択肢は5つで、その中に必ず「わからない」という選択肢を含んでいる。これは当て推量による正解を排除するための措置である

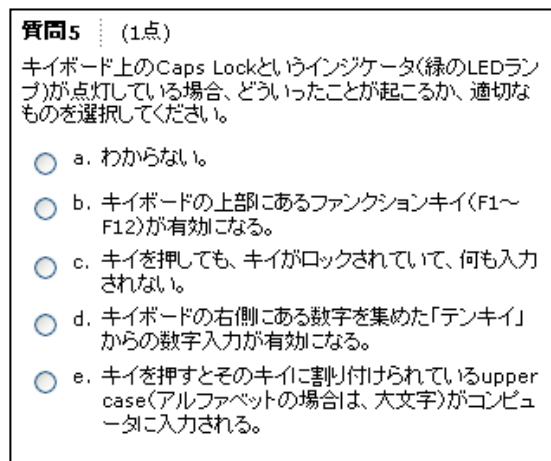


図3 問題項目の例（Caps Lock）

本来、利用能力は利用するコンピュータの機種によらない形で測定するのが理想だが、今回の測定の目的は、大学における勉学で支障が無いところまで利用能力を獲得してもらう指針とするためなので、本学のコンピュータ演習室で利用できる Windows 基本ソフトに特化する形で問題項目の作成を行なった。

2004～2007年度の4年間の p テスト、a テストにおける学生の能力（スコア）分布を、図4に示す。図中で2004～2006の後につくアルファベットは、そのデータが p テスト（プレースメントテスト）か、a テスト（アチーブメントテスト）かを示している。

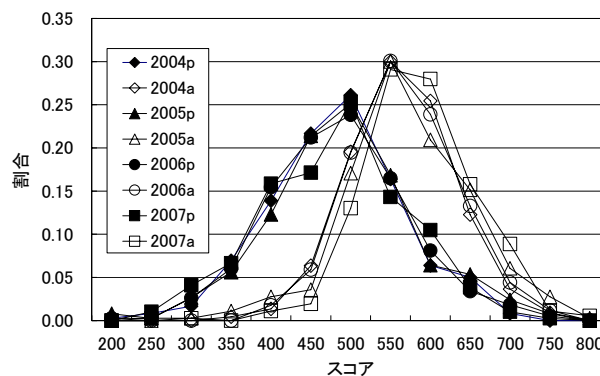


図4 2004-2007 スコア分布

この図より以下のことが示唆される[7][8][9][10].

- (1) 項目応答理論を使ったスコアにより、学生のコンピュータ利用能力が安定的に測定されている。
- (2) 2006年度以降に入学した学生は、高校で教科「情報」を学んできているが、ここで調査している基礎的なコンピュータ利用能力の範囲内では、それ以前の学生と違いが認められない。
- (3) 2006年度以降は利用教育に習熟度別クラスを導入したが、以前のクラス編成の場合と利用教育終了後のスコア分布では違いがない。

(4) IRTにより学生の基礎的なコンピュータ利用能力を安定的に測定することができる。

ただし、ここで測定している「基礎的なコンピュータ利用能力」とは、一般的な意味ではなく、表1で示した問題項目で測定できる範囲内のものであるということを注意されたい。

4. 習熟度別クラス

2006, 2007年度は、利用能力に応じたリメディアル教育の前段階として、pテストの結果をもとに、クラスをほぼ半分に分割し、基本クラス（以下、Bクラスと略す）とアドバンスクラス（以下、Aクラスと略す）という習熟度別クラス編成で利用教育を行った。ただし、今回の習熟度別クラス編成は、同一科目という枠内で行ったため、習熟度により教育内容を大きく変えるということはできず、習熟度が同程度の学生を同じクラスに所属させ、利用教育を円滑に進めようというものであった。同一科目という制約のため、BクラスもAクラスも課題内容や評価項目は同一にした。

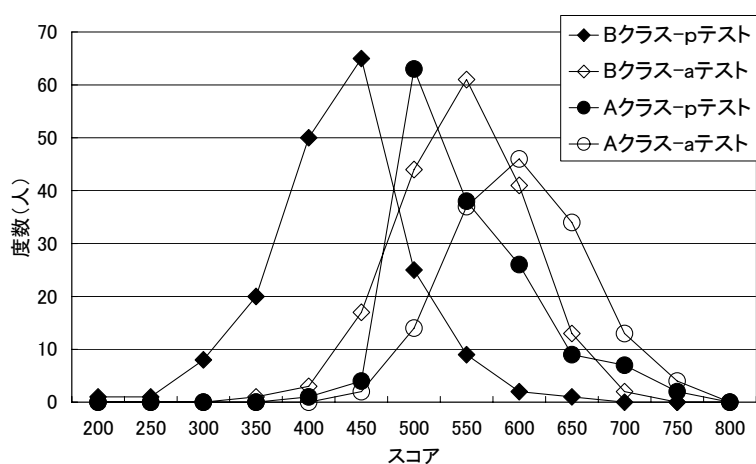


図5 クラス分類ごとのスコア分布（2006年度）

さて、2006年度の習熟度別クラスの場合をもう少し詳しく見てみる。2006年度には、Bクラス（基本クラス）を5クラス、Aクラス（アドバンスクラス）を4クラス開講した。クラスの分類ごとにまとめたpテストとaテストのスコア分布を図5に示す。pテストの分布がクリアカットでないのは、学生全体をまとめてBクラスとAクラスに分割をしたのではなく、学部または学科ごとに分割したため、分割の基準が多少変動したためである。この図より、Bクラスの学生の方がスコア分布の移動が大きい傾向が見て取れる。この点を確認するため、各学生のpテストのスコアと、aテストとpテストのスコアの差をプロットした散布図を図6に示す。

この図より以下のことが示される[7][8].

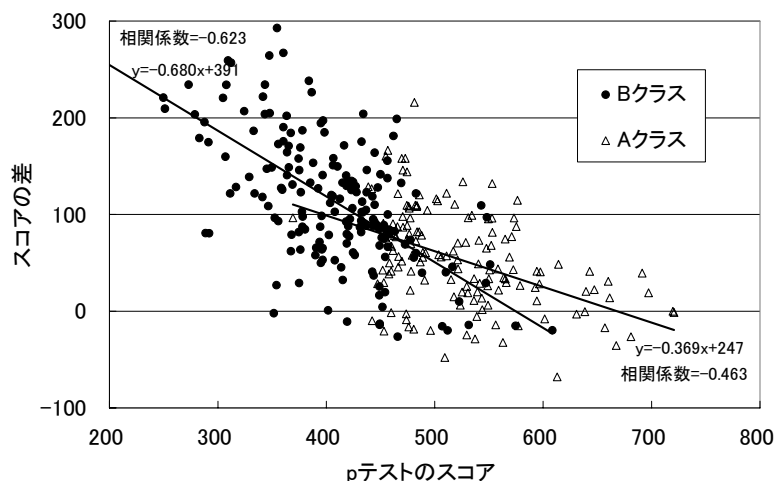


図 6 スコア上昇の散布図 (2006 年度)

- (1) 全体的な傾向としては、p テストのスコアが低い学生の方が、利用教育後のスコアの伸びが大きい。
- (2) B クラスと A クラスの回帰係数はそれぞれ、 -0.680 と -0.369 であり、B クラスの方がスコアの伸び率が大きい。

習熟度別クラス編成を行った結果、習熟度が高いグループの方が、能力の伸びが低いという結果になった。これは、今回の習熟度別クラス分けは、同一の科目の中で試験的に実施したので、不公平が無い様に課題や評価項目を同一にしたためであろうと考えられる。A クラスでは、授業中に高度な説明を行うことはあったが、基本的な教授項目は B クラスと同じであり、A クラスの学生にとっては若干物足りなかったのではないかと推測される。習熟度別クラス編成の利点として、A クラス、B クラスとも、授業の進行はスムーズになったという感想が、教育担当者からあがっている。

5. 学習モジュールの「知識度」

本研究の目的は、p テストで推定されたスコアの学生に対して、どの教材を提供するかを判断を行う指標を開発することである。そのため、2007 年度の利用教育の中で、各学習モジュール（利用教育中の教授単位：パワーポイントのファイルという形式でモジュール化されている）を学習した直後に、その学習内容を知っていたかどうかのアンケートを行った。表 2 に学習モジュールとして用意した 28 個のタイトルとその分類を示す。アンケートは、「知っていた」から「知らなかった」までを 4 件法で回答を求めたが、そこから「知っていた」と「知らなかった」に 2 値化し、その結果から IRT の 2 母数ロジスティックモデルを使ってモデル化する。IRT による推定は BILOG を使って、各学習モジュールに対応する ICC を求めた。もちろん、回答者のスコアは、前述の p テストで推定されたものを利用した。通常の ICC の場合、縦軸はその項目に対する「正答率」になるが、今回は、利用教育を受ける前に、その学習モジュールの内容を知っている確率（以下「既知確率」と呼ぶ）となっている。

図 7 に、各スコアの学生がいくつかの学習モジュールに対してどの程度の既知確率を持っていたかの実測値（アンケート結果から計算した割合）と、IRT の 2 母数ロジスティックモデルにより推定した既知確率の対応を示す。

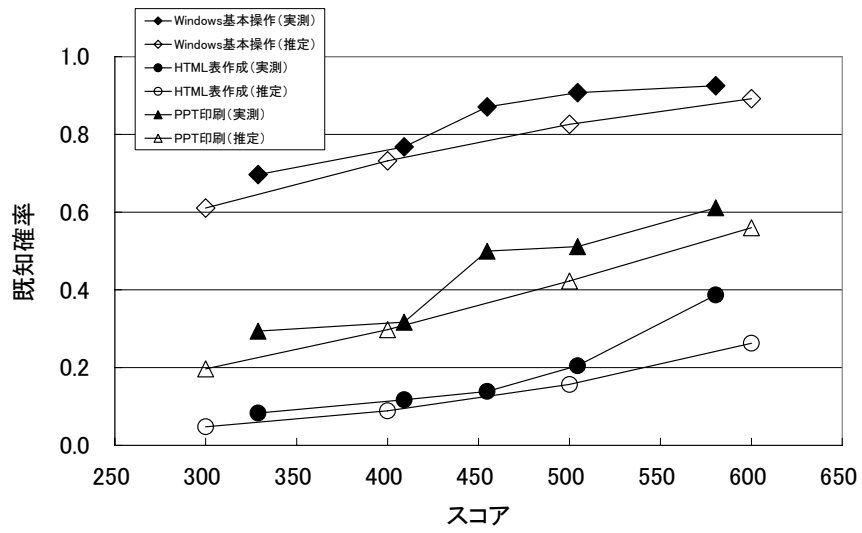


図7 実測値と推定値の対応

表2 学習モジュールのタイトル

分類	学習モジュールタイトル
Windows 基礎	パソコン基本操作
	Windows 基本操作
	Windows MO の使い方
日本語入力	Word の起動と終了
	日本語入力 基礎
	日本語入力 言語バー
	日本語入力 文章の変換
電子メール	電子メール
ワープロ	Word 文章の編集
	Word 書式（フォントと段落）の操作
	Word 印刷
	Word 数式入力
ファイル操作	ファイル操作1（保存、複製、削除など）
	ファイル操作2（検索、ファイル属性など）
PowerPoint	PPT テキストの編集
	PPT 図表の挿入
	PPT アニメーション
	PPT スライドショー
	PPT 印刷
HTML	HTML はじめの一步
	HTML 簡単なページの作成
	HTML 文字飾り
	HTML 画像の表示
	HTML 写真の取込
	HTML ハイパーリンク
	HTML 表の作成
	HTML トラブルシューティング
	ホームページビルダ

IRT による推定が、過小評価気味ではあるが、比較的良い対応をしていることがわかる。2 母数ロジスティックモデルの困難度パラメータは、この場合は、該当モジュールの学習内容を、利用教育前に知っている確率が 50%に達する学生のスコアに相当するので、ここでは、「困難度」ではなく「知識度」[11][12]と呼ぶことにする。各学習モジュールの知識度と識別力（従来の識別力パラメータと同じ）の散布図を図 8 に示す。

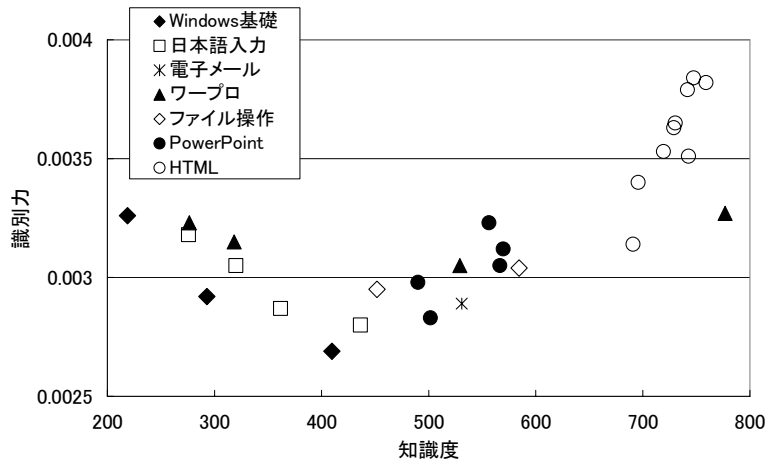


図 8 各学習モジュールの知識度

各々の点は各モジュールの知識度と識別力を示しているが、点の種類（●，▲，◆など）は、学習モジュールの分類（日本語入力，ワープロ，HTML など）でまとめた。各モジュールの知識度は、大分類ごとに大体かたまっているが、ワープロなどで一つ知識度の非常に高いものがある。これは、「Word 数式入力」に『対応しており、初等中等教育でほとんど学んで来なかった項目であることがわかる。他の学習モジュールの知識度も担当教員の実感とおおしく外れたものはない。

6. スキップ確率

ロジスティック関数を用いた学生の学習モジュールの知識度モデルを使って、利用教育の中で学生がどの程度各学習モジュールをスキップできるかを推定する。図 9 は、2007 年度の学生の p テストのスコアから、知識度が 200~600 までのモジュール（識別力は 0.003 に固定）の学習内容の既知確率が 0.5~0.9 を超える学生数の割合をグラフ化したものである。今回の学習モジュールの知識度は、図 8 に示したように Windows 基礎関連が 214~408、日本語入力関連が 276~436、電子メールが 531、ワープロが 276~777、ファイル操作が 452~584、PPT が 490~570、HTML が 691~759 であった。たとえば、あるモジュールの既知確率が 0.7 以上となるスコアを p テストで取得した学生は、そのモジュールを学習しなくてよいということにすると、Windows 基礎関連で、最も知識度が低いモジュール（知識度 214）の場合、8 割の学生は学習しなくてもよいことになり、最も知識度が高いモジュール（知識度 408）の場合、1 割程度の学生がスキップできることになる。また、高校までに学習する機会がほとんどない HTML 関連のモジュールは、知識度が最も低いもの（知識度 691）でも、ほぼ 100%の学生が学ぶ必要があることになる。この例では、学習モジュールをスキップできるかどうかの既知確率の基準値を 0.7 としたが、この数値に特に理由はない。学習モジュールをスキップするために、どの数値が適切であるかは、今後、研究を継続していく必要がある。

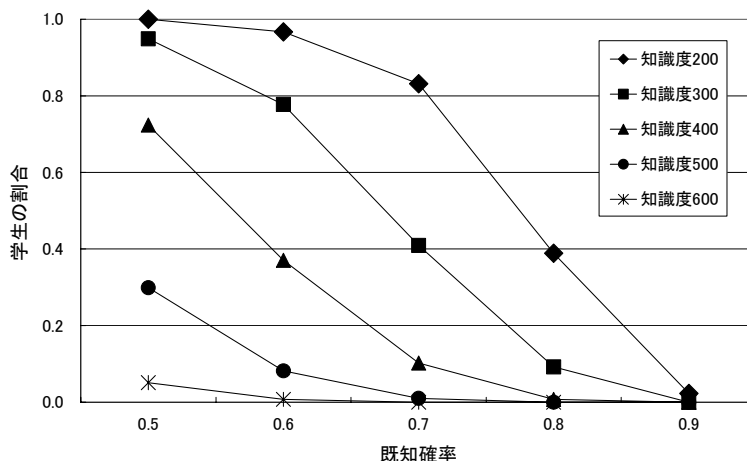


図9 ある既知確率を超える学生の割合

7. 本手法における課題

IRTのスコアに対応して学習モジュールを選択するための基準をある程度示すことができたが、この手法の課題として以下のものが考えられる。

(1) 既知確率の基準値

基準値を増減させることにより、モジュールをスキップできる学生数は増減する。知っていない学生をスキップさせないことを目的にするならば、基準値をあげればよいが、そうすると、すでに知っているにもかかわらずモジュールを学習しなくてはならない学生が増加する。逆に、基準値を下げると、知らないのにモジュールを学習せずに過ぎてしまう学生が増加する。適切な既知確率の基準値は、今後の研究の中で決めていかななくてはならない。

(2) この手法による判断と現実とのギャップ

この手法では、各モジュールの学習内容が個々の学生に習得されているかどうかを細かく見ているわけではないので、スコアから判断した学習すべきモジュールと、本当にその学生に必要な学習内容との間にはギャップが存在する。それゆえ、実際の学習の際には、そのギャップを埋めるための方策が必要になる。

(3) インストラクショナルデザイン手法との比較

各モジュールの学習内容が習得されているかどうかをテスト等により一つ一つチェックし、習得していない場合のみ、そのモジュールを学習するという手法も世の中には存在する。インストラクショナルデザインの手法がこれに相当するが、この手法は、教員側からすると、各学習モジュールの関係（スキルマップと呼ばれる）をトップダウンに明らかにし、それに従って学習コンテンツの作成が必要になり、かなり手間がかかる。また、学生側からしても、すべてのモジュールに関するテストを受けなくてはならないので、学習内容が多くなると、膨大な手間が発生する。その点、本手法は比較的簡単なテストとアンケートから、必要な学習モジュールの選択が可能になることがメリットであると考えられる。ただしその場合でも、課題(2)のギャップを埋めるために、学習モジュールをスキップさせる前に、何らかのテストが必要になる可能性は大きい。この場合、本手法で大まかに、各学生が学習すべきモジュールを特定し、スキップしてもよいと判断されたモジュールのみ、テストを実施するという、本手法と個々の項目のテストを行うという手法のハイブリッド的な手法を開発する必要がある。

8. 結論

IRT の 2 母数ロジスティックモデルを使い、学習モジュールのスコアと既知確率の関係をモデル化できた。この手法は、正確さには課題があるが、実用的という点では、他の手法に比べてすぐれている。本研究で現在までに以下のことが確認できた。

- (1) 図 3 でみるように、このモデルは若干過小推定になっているが、大まかなところでは既知確率の妥当な推定と考えられる。
- (2) 各学習モジュールの知識度（既知確率が 0.5 に相当するスコアの値）は、担当教員の実感と大きく乖離はない。
- (3) 2007 年のデータより、この方式を使った場合、現在提供している学習モジュールをスキップできる学生の割合を求めた。非常に知識度が低いもの（200）で 8 割、そのつぎに知識度が低いもの（300）で 4 割、中程度の知識度（400）で 1 割程度の学生がスキップできる可能性が示されたが、中程度より知識度が大きい学習モジュールに関しては、ほとんどスキップできる学生はいないことがわかった。（ただし、既知確率の基準値を 0.7 とした場合）

今後は、課題のところでも述べたように、実用化に向けていくつかの点を解決していかなくてはならないと考えている。

[参考文献]

- [1] 川合慧 監修, 河村一樹 編著, 情報とコンピューティング. オーム社, 東京, 2004.
- [2] 山川修, 菊沢正裕, 大学における情報基礎教育カリキュラムの実践的研究. 日本教育工学会論文誌, 30 : 231-238, 2006.
- [3] Lord, F. M., *A theory of test scores*. Psychometric Monograph, No.7, 1953.
- [4] 芝祐順 編, 項目反応理論. 東京大学出版会, 東京, 1991.
- [5] Lord, F. M., & Novick, M. R., *Statistical theories of mental test scores*. Reading, MA : Addison-Wesley, 1968.
- [6] 芝祐順, 語彙理解尺度作成の試み, 東京大学教育学部紀要, 17, 47-58, 1978.
- [7] Osamu Yamakawa, Takeyuki Tanaka, Masahiro Kikusawa, *Effective measurement of classes based on ability by using the item response theory*, Program and abstracts of 8th International Conference on Information Technology Based Higher Education and Training (ITHET2007), p.87, 2007.
- [8] Osamu Yamakawa, Masahiro Kikusawa, Takeyuki Tanaka, *Ability Measurement in Computer Literacy using the Item Response Theory for Adaptive Learning*, Proceedings of E-Learn 2007 (World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education), pp.549-554, 2007.
- [9] 山川修, 田中武之, 菊沢正裕, 「項目応答理論を使った新入生の情報基礎能力の継続的な分析」, 情報処理学会研究会報告 (第 2 回 CMS 研究会), pp.45-48, 2006.
- [10] 山川修, 田中武之, 菊沢正裕, 「項目応答理論を利用した能力別編成クラスの効果測定」, 日本教育工学会第 22 回全国大会講演論文集, pp.975-976, 2006.
- [11] 山川修, 菊沢正裕, 田中武之, 「項目応答理論を使った学生の能力推定に対応する教材選択手法の

開発」, 第4回 WebCT 研究会予稿集, pp.39-42, 2007.

- [12] Osamu Yamakawa, Masahiro Kikusawa, Takeyuki Tanaka, *KNOWLEDGE DEGREE OF LEARNING MODULES CORRESPONDING TO THE IRT SCORE*, Proceedings of the Seventh IASTED International Conference WEB-BASED EDUCATION (WBE2008), in Press.