

問題の出題パターンに着目した難易度推定方法に関する検討

池田信一^{†1} 高木輝彦^{†2} 高木正則^{†3} 勅使河原可海^{†1}

近年、web上で実施するテストである「e テスティング」が注目されている。e テスティングでは問題の難易度などの統計データを管理するアイテムバンクを用意することで、より幅広いテストの出題が可能となる。問題の難易度の定量化には、一般的に項目反応理論 (IRT) が用いられ、IRT で問題の難易度を推定するためには、予め多くの学習者に問題を解答させる必要がある。しかし、アイテムバンク内のすべての問題を解答させることは困難であるため、解答履歴のない問題の難易度の推定が課題となる。そこで、我々は解答履歴のない問題の難易度の推定を目的とし、問題の出題パターンに着目した難易度の推定方法を検討する。問題の難易度は問題内で出現する知識の問われ方によって変化すると考えられる。そのため、まず問題の出題パターンと難易度の関係性を分析した。そして、分析結果から、出題パターンに着目し難易度を推定する手順を検討した。IRT で推定した難易度との比較実験の結果、提案手法により IRT で推定した難易度と近い値で推定できることが分かった。

A Study on a Method of Estimating the Difficulty of Quizzes Focused on Quiz Types

Shinichi IKEDA^{†1} Teruhiko TAKAGI^{†2}
Masanori TAKAGI^{†3} Yoshimi TESHIGAWARA^{†1}

In recent years, tests which carried out on the web - “e-testing” have been attracting much attention. In e-testing, wider range of tests can be performed by preparing item bank managed as a statistical data, such as based on the difficulty level of quizzes. In general, Item Response Theory (IRT) is used to quantify the difficulty level of quizzes. In order to estimate the difficulty of the quizzes, students need to answer the quizzes in advance; However, it is hard to make students answer all quizzes in the item bank.

^{†1}創価大学大学院工学研究科

Graduate School of Engineering, Soka University

^{†2}電気通信大学大学院情報システム学研究科

Graduate School of Information Systems, The University of Electro-Communications

^{†3}岩手県立大学ソフトウェア情報学部

Faculty of Software and Information Science, Iwate Prefectural University

Therefore, a method of estimating the difficulty level of not answered quizzes is studied. The study estimate the difficulty level of not answered quizzes by focusing on quiz type. The difficulty level of the quizzes could be changed depending on how knowledge is asked. Thus, firstly the relationship between quiz types and difficulty level are analyzed. Next, from the result of the analyses, a procedure for estimating the difficulty level of quizzes focused on quiz types is discussed. The experiment results of comparing the proposed method with an IRT method showed that the value of the difficulty level estimated using IRT is close to the value estimated using the proposed method.

1. はじめに

近年、Web上で実施するテストである「e テスティング」が注目されている[1]。e テスティングでは、問題ごとの細分化された教育目標や、問題の難易度などのテスト実施後の統計データを管理するアイテムバンク (item bank) [2]を構築することで、より精度の高いテストやより学習効果のあるテストを実施することが可能となる。そのため、e テスティングでは、アイテムバンクを充実させることが重要となる。

問題の難易度の定量化には、一般的に項目反応理論 (Item Response Theory: IRT) [3] が用いられる。IRTで問題の難易度を推定するためには、予め多くの学習者に問題を解答させる必要がある。しかし、システムの規模によってはアイテムバンク内に膨大な量の問題が存在するため、それら全てを解答させることは困難である。そのため、アイテムバンク内に存在する解答履歴のない問題の難易度の推定が課題となる。

そこで、本研究では、解答履歴のない問題の難易度の推定を目的とし、問題の出題パターンに着目した難易度推定方法を提案する。問題の難易度は問題内で出現する知識の問われ方によって変化し、また、類似した問題同士の難易度には関係性があると考えられる。そのため、類似問題において、知識の問われ方の変化に伴う難易度の差に着目することで、解答履歴のない問題の難易度が推定可能であると考えられる。具体的には、IRT で複数の問題の難易度を推定することで、類似問題の難易度を出題パターンに着目し推定する手法を検討する。なお、本研究では多肢選択式の問題を対象とする。

以下、2章では、IRTを用いた難易度の推定について説明する。3章では、問題の出題パターンと類似問題を定義し、研究目標や難易度を推定する際の課題を明確にする。4章では、それらの課題を解決するために分析を行い、その分析結果を基に独自の難易度算出式を用いた難易度推定手法を提案する。さらに、5章では、IRTで推定した難易度と、提案手法で推定した難易度の比較実験を行い、提案手法による難易度推定の妥当性の検証を行う。6章では、実験結果から明らかになった提案手法の問題点についての分析を行い、7章で今後の課題について述べ、最後に、8章でまとめる。

2. 項目反応理論

本研究では、予め IRT を用いて複数の問題の難易度を推定することで、それらの類似問題の難易度を推定する手法を検討している。IRT は、IRT モデルと呼ばれる統計モデルを用いて、各問題の統計的な性質を明らかにする。問題の性質は、縦軸に正答確率を、横軸には受験者集団に依存しない被験者の学力を表す特性値 θ によって描かれた項目特性曲線 (Item Characteristic Curve: ICC) で表される。

学習者の能力特性を表す能力母数が θ である学習者 i の問題 j への正答確率 $P_j(\theta_i)$ は、以下の式(1)により定義される。

なお、本研究ではもっとも一般的に利用されている 2 パラメータ・ロジスティックモデル (2PLM) を適用する。

$$P_j(\theta_i | a, b) = \frac{1}{1 + \exp\{-Da_j(\theta_i - b_j)\}} \quad (1)$$

ここで、 D は $D=1.7$ となる定数である。 a は識別力母数であり、問題 j が学習者の能力を識別する力を表す。数学的には $-\infty$ から ∞ の範囲で定義されるが、識別力が負であると $P(\theta)$ は単調減少関数になってしまい、その場合、特性値 θ の値が大きい学習者ほど正答確率が低くなってしまふ。そのような問題は問題自身に欠陥があるとして、分析を行う前に削除するため、実質的に識別力はせいぜいの値だけを扱うものとなる。 b は難易度母数であり、難易度の値が大きいほど難しい問題となる。難易度の値も数学的には $-\infty$ から ∞ の範囲で定義される。ただし、特性値 θ の分布が平均 0 分散 1 に標準化されているとき、難易度はおよそ -3.0 から 3.0 の間で推定される[4]。

図 1 に $P(\theta|0.5,0)$, $P(\theta|1.0,1)$, $P(\theta|2.0,0)$ とした場合の ICC を示す。ここで、 $b=\theta$ に着目すると、正答確率 $p=0.5$ となる。このことから、難易度はその問題を五分五分で正解できる能力レベルを表していることが分かり、グラフが右に行くほど難易度が高くなる。また、識別力が大きい問題ほど ICC の傾きが急になる。2PLM では、 $P(\theta)=0.5$ のときに傾きは最大となり、そのときの傾きの値が識別力 a である。識別力の高い問題は $b=\theta$ 付近で大きく正答確率が変化し、逆に識別力の低い問題では正答確率に大きな変化がない。つまり、識別力とは、その問題の難易度付近の能力特性を持つ学習者を識別する力のことを表している。

さらに、被験者 i の n 個の項目に対する解答を $\mathbf{u}_i = \{u_{i1}, u_{i2}, \dots, u_{ij}, \dots, u_{in}\}$ としたとき、局所独立の仮定から、ベクトル \mathbf{u}_i となる確率は

$$P_j(\mathbf{u}_i | \theta_i) = \prod_{j=1}^n p_j(\theta_i)^{u_{ij}} q_j(\theta_i)^{1-u_{ij}} \quad (2)$$

と表現される。ここで、 $q_j(\theta_i)$ は被験者 i が項目 j に誤答する確率である。

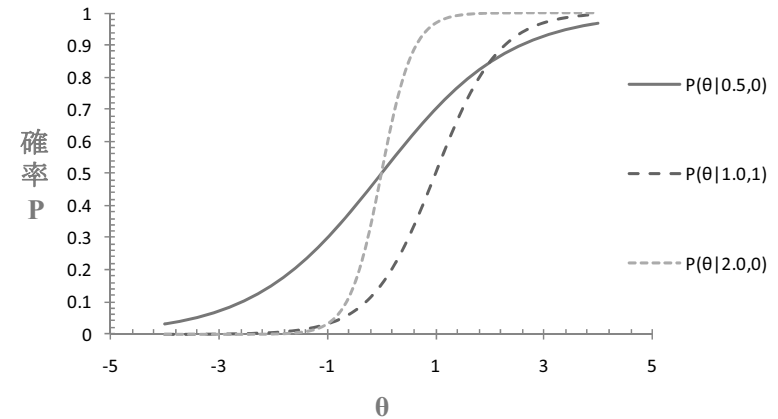


図 1 2PLM の項目特性曲線

Figure 1 Item Characteristic Curve of 2PLM

このモデルから、学習者のテストへの解答を基に学習者の能力母数 θ 、問題の識別度母数 a と難易度母数 b を同時最尤推定法やベイズ推定法などを用いて推定することが可能である[1]。一方で、これらの母数をより正確に推定するためには、予め多くの学習者に問題を解答させる必要がある。

3. 研究目標と検討課題

3.1 研究目標

本研究では、問題の出題パターンと難易度の関係性を分析し、その分析結果から解答履歴のない問題の難易度の推定について検討する。具体的には、予め IRT で複数の問題の難易度を推定しておき、それらの類似問題の難易度を出題パターンに着目し推定する手法を検討する。そのため、IRT で推定した難易度が基準となるので、本研究で推定した難易度と IRT で推定した難易度の比較を行い、IRT で推定した難易度と近い値を目指す。また、一回の推定でより多くの問題の難易度の推定を目指す。

3.2 問題の出題パターン

難易度推定法の検討をするためには、まず、問題をいくつかの出題パターンに分類する必要がある。

我々は先行研究において、知識の問われ方に基づき問題を 11 の出題パターンに分類した[5]。問題の出題パターンは、例や説明、種類や属性、関連する事柄などが問われているのか、また、正しいものを選ぶのか (+)、誤っているものを選ぶのか (-) に

よって細かく分類を行った。表 1 に、問題を出題パターンで分類した結果とそれらの例を示す。「その他」というのは、計算問題やある事柄の順序問題、穴埋め問題、または、不備のある問題などの問題である。

表 1 問題の出題パターン

Table 1 Quiz types

パターン ID	出題パターンと例
Pa+	ある専門用語について正しい例や説明などを選択する問題 例：モル信号の説明として正しいものを選び。
Pa-	ある専門用語について誤った例や説明などを選択する問題 例：リアルタイム処理分散システムの例として、適切でないものを選び。
Pb+	ある専門用語についてその種類・属性と同じ専門用語を選択する問題 例：以下の中で目視的コミュニケーションはどれか。
Pb-	ある専門用語についてその種類・属性と異なる専門用語を選択する問題 例：次のうち目視的コミュニケーションでないものを選び。
Pc+	ある専門用語に関連する事柄について正しい例や説明などを選択する問題 例：分散環境におけるデータ管理に関する記述として、適切なものはどれか。
Pc-	ある専門用語に関連する事柄について誤った例や説明などを選択する問題 例：コンピュータネットワークの構築理由として誤っているものを示せ。
Pd+	ある専門用語に関連する事柄について関係する専門用語を選択する問題 例：電話通信をするために、適しているものはどれか。
Pd-	ある専門用語に関連する事柄について関係しない専門用語を選択する問題 例：コンピュータネットワークの応用例として、銀行やコンビニエンスストアにないものはどれか。
Pe+	各選択肢に記述された専門用語とその説明の組み合わせのうち正しい組み合わせを選択する問題 例：OSI 参照モデルのうちの 4 つの階層で正しい説明がされているものを選び。
Pe-	各選択肢に記述された専門用語とその説明の組み合わせのうち誤った組み合わせを選択する問題 例：以下に示した語句の説明の中で、最も正しくないものはどれか。
Pf	ある専門用語についての例や説明からその専門用語を選択する問題 例：単一のホストコンピュータを複数の端末から利用するホスト集中システムとは。
	その他（計算、穴埋め、不備）

3.3 類似問題の定義

問題の難易度は、問題の分野や問題内で問われている知識によって異なると考えられる。そのため、出題パターンと難易度との関係性を分析するためには、同じ知識について問われている問題同士を比較する必要がある。そこで、本研究では類似問題を「問題で問われている知識や解決の中心となる知識が一致する問題」と定義する。類似問題間で難易度の比較を行うことで、同じ知識に対しての相対的な難易度の変化の分析が可能となる。

また、問題の管理方法として、問題の類似性に着目し、問題をコンピュータで自動的に管理する手法が提案されている[6]。本研究では、この手法により問題が類似問題群に分類されているものとする。以下に本研究で定義する類似問題の例を示す。

<問題 1>

利用者のパソコンから電子メールを送信するときや、メールサーバ間で電子メールを転送するとき転送されるプロトコルはどれか。

- | | | | |
|---------|---------|--------|---------|
| 1. SMTP | 2. HTTP | 3. POP | 4. IMAP |
|---------|---------|--------|---------|

<問題 2>

SMTP の説明で間違っているものはどれか。

- このプロトコルに従って、送信元のコンピュータから宛先のユーザボックスがあるコンピュータまで転送される。
- サーバ - サーバ型のプロトコル。
- MTA はメール転送を受けつけるときには、SMTP のサーバ、メールを他のサーバに転送する場合はクライアントとして振る舞う。
- クライアントが送信元のアドレスを直接サーバに伝えている。クライアントが自由に自分のメールアドレスを変更できる。

両問題共に作問者が問っている知識は SMTP である。問題 1 では SMTP の説明を問題文で述べ、選択肢の中から SMTP を選択させる問題であり、問題 2 では、SMTP の説明として不適切な選択肢を選ばせる問題である。問題 2 では適切な選択肢を選ばせるように問題を作成することもできる。このように、1 つ、あるいは、複数の知識を中心として様々な問題を作成できるため、この知識に着目して類似問題を定義付けた。また、知識とは分野特有の概念や法則、人物、歴史など様々であり、ほとんどが専門用語である。

3.4 検討課題

問題の難易度は問題内で出現する知識の問われ方によって変化すると考えられ、また、類似問題間で難易度と知識の問われ方に関係性があると考えられる。例えば、専門用語について問う問題であっても、説明文から正しい専門用語を選択させる問題と、専門用語の正しい説明文を問う問題では難易度が異なると考えられる。また、正しいものを選ぶのか、誤っているものを選ぶのかでも変化すると考えられる。しかし、このような知識の問われ方の違いによってどの程度難易度が変化するのは明らかになっていない。

そこで、本研究では、アイテムバンク内の難易度が推定済みの問題を基に、それぞれの出題パターンの難易度の差を分析し、その結果から独自の難易度算出式を検討する。

4. 出題パターンに着目した難易度推定の検討

4.1 出題パターンと難易度の分析

難易度推定の検討をするにあたって、まず、出題パターンごとの難易度の差を分析した。なお、今回は対象とする問題数の関係上、(+)と(-)の分類は行わず、Pa~Pfの6つの出題パターンについて分析を行う。分析の手順を図2に示す。まず、IRTを用いて複数の問題の難易度を推定する(図2中①)。対象とする問題は、本学で開講されている講義「コンピュータネットワーク論」において、高木らにより開発された作問システムCollabTest[7]を用いて、'06年度から'08年度までに作成された問題とし、各年度約80名の解答履歴をもとに難易度の推定を行った。次に、難易度を推定した問題ごとに類似問題を検索する(図2中②)。さらに、検索した類似問題の中から難易度が推定してある問題を抽出し、その問題の難易度とIRTで推定した難易度の差を算出する(図2中③)。問題の難易度は問われている知識によって異なると考えられるため、問題の比較はすべて類似問題間で行い、相対的な難易度の差を分析する。最後に、出題パターンの組み合わせごとに、難易度の差の値を集計する(図2中④)。表2に、図2の手順により出題パターンごとに集計した結果を示す。

表2の結果から、類似問題間において出題パターンが変化することで、難易度に差が出るのがわかる。また、出題パターンごとの難易度の差をみるとPdやPcは他の出題パターンに比べ難易度が高めに推定され、逆にPaやPeは低めに推定されている。これは、PdやPcは専門用語に関する事柄を問う問題であり、知識を応用して使う必要があるため難易度が高くなっていると考えられる。一方で、PaやPeは知識を応用する必要はなく、その知識を知っていれば解けるため、難易度が低くなったと考えられる。

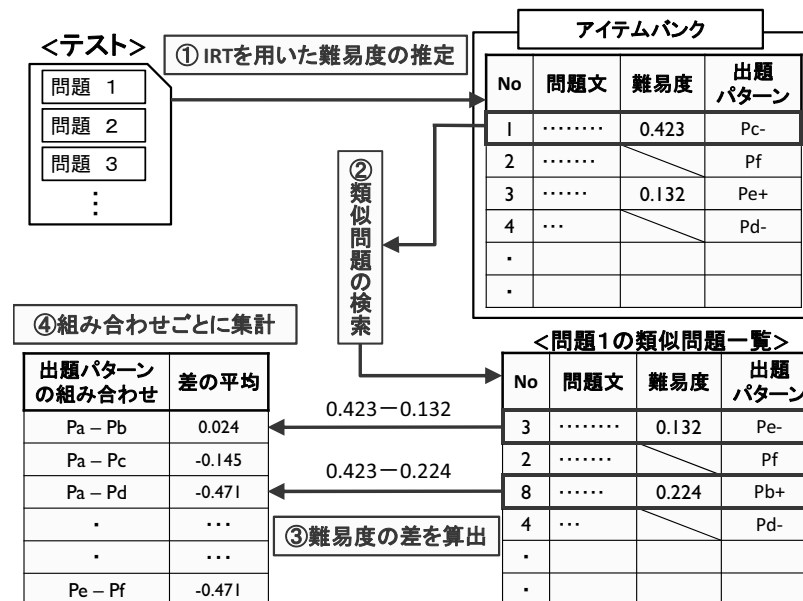


図2 出題パターンごとの難易度比較手順

Figure 2 Difficulty comparison procedure of each quiz types

また、出題パターンの組み合わせごとの難易度の差の平均値を表3に示す。なお、表3の値は行の出題パターンの難易度から列の出題パターンの難易度を引いた値の平均値となっている。

この表では、出題パターンがPaの問題は、出題パターンがPc問題に比べて難易度が約0.14465低いことになる。さらに、PaとPbの組み合わせではほとんど難易度が変わらないのに対し、PaとPf、また、PbとPfの組み合わせをみるとPbとPfの方が低く推定されているのが分かる。このことから、出題パターンの組み合わせによって難易度の差が変化すると考えられる。これらの分析結果を基に、独自の難易度推定手法を提案する。

表 2 出題パターンごとの難易度の差の平均値

Table 2 The average difference in the difficulty of each quiz types

出題パターン	難易度の差の平均
Pa	-0.13758
Pb	-0.12756
Pc	0.182587
Pd	0.442029
Pe	-0.22627
Pf	0.009115

表 3 出題パターンの組み合わせごとの難易度の差の平均値

Table 3 The average difference in the difficulty of each combination of quiz types

出題パターン	Pa	Pb	Pc	Pd	Pe	Pf
Pa		0.024	-0.145	-0.471	0.248	-0.018
Pb	-0.024		0.022	-0.535		-0.333
Pc	0.145	-0.022		-0.384	0.261	0.311
Pd	0.471	0.535	0.384		0.721	0.291
Pe	-0.248		-0.261	-0.721		-0.539
Pf	0.018	0.333	-0.311	-0.291	0.539	

4.2 出題パターンに基づく難易度推定手順

図 3 に本研究で提案する難易度算出手順を示す。まず、予め、アイテムバンク内の複数の問題の難易度を、IRT を用いて推定する (図 3 中①)。次に、解答履歴のない問題 (以下、推定問題) の類似問題を検索し、その中から難易度が既知である問題 (以下、比較問題) を抽出する (図 3 中②)。そして、推定問題と比較問題の出題パターンを比較し、2 つの出題パターンの組み合わせにおける難易度の変化量を求める (図 3 中③)。なお、難易度の変化量は、①で難易度を推定した問題の中から、類似問題間で難易度が比較可能な問題の難易度の差を算出し、その結果を基に設定する。最後に、①で推定した比較問題の難易度に、③で求めた推定問題と比較問題の出題パターンの組み合わせにおける、難易度の変化量を足すことで推定問題の難易度を算出する。また、式 (3) に本研究で提案する難易度算出式を示す。

$$PD = \frac{1}{n} \sum_{j=1}^n (sw_{ij} + d_j) \quad (3)$$

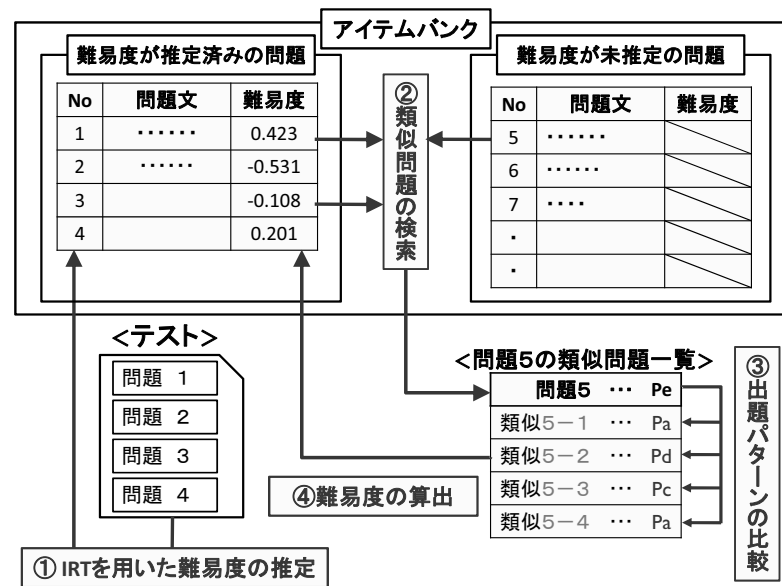


図 3 難易度算出手順

Figure 3 Calculation procedure of difficulty

ここで、 PD は n 問の比較問題 j から算出した推定問題 i の難易度である。また、 w_{ij} は出題パターンの組み合わせの重みを表し、 s はその重み 1 分の変化量を表す定数である。具体的な値は、前節の分析結果を基に設定する。さらに、 d_j は IRT により推定した比較問題 j の難易度を表している。

難易度推定の流れは、推定問題 i と比較問題 j の出題パターンの組み合わせの重み w_{ij} を基に、難易度の変化量 sw_{ij} を d_j に足すことで比較問題 j に対する難易度を算出する。同じ処理を残りの $n-1$ 個の比較問題に対して行う。最終的にその平均値を推定問題 i の難易度とする。

同じ処理を、類似問題群他の問題でも行う。類似問題群ごとの難易度算出のフローチャートを図 4 に示す。提案手法により難易度を推定した問題も、他の問題の比較問題になるものとするため、提案手法で繰り返し難易度の推定を行うと、比較問題の数が増える。従って、提案手法では繰り返し難易度の推定を行い、前回の推定値との差が予め設定した閾値より小さくなったところで、その問題の推定を終了する。

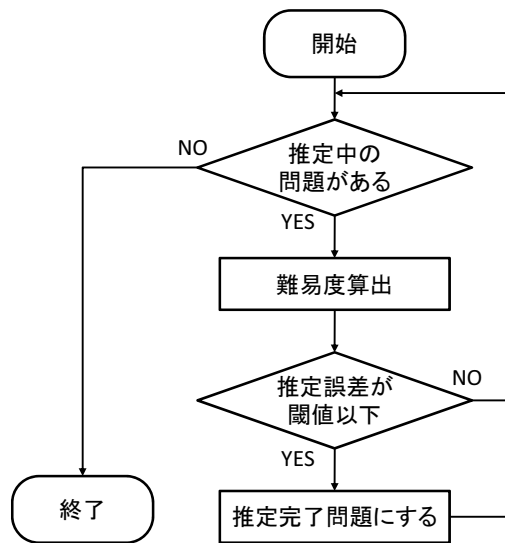


図 4 類似問題群ごとの難易度算出のフローチャート
Figure 4 Flowchart for calculating the difficulty of each similar quiz group

4.3 期待される効果

本方式では、予めIRTで複数の問題の難易度を推定することで、それらの類似問題の難易度を学習者に解答させることなく推定することが可能となる。そのため、アイテムバンク内に難易度が既知である問題が増えれば、それらの問題を効果的に利用することができる。まず、学習者の理解度に適した適応的なテスト出題が挙げられる。図 5 に類似問題群からの適応的なテスト出題方式を示す。学生がテストを要求すると（図 5 中①）、予め蓄積されている類似問題群から各問題の類似問題を自動で検索し（図 5 中②）、それらの問題の中から学生の理解度に適した問題を提示する（図 5 中③）。一般的な適応型テストでは、類似問題を繰り返し解くことはできないが、この方式では、テストを構成する各問題に類似問題が登録されている場合、毎回異なる類似問題が動的に出題されるように構成することで、テストを繰り返し解答することによる答えの丸暗記の防止や、多角的な学習により、より深い理解の定着が期待できる。

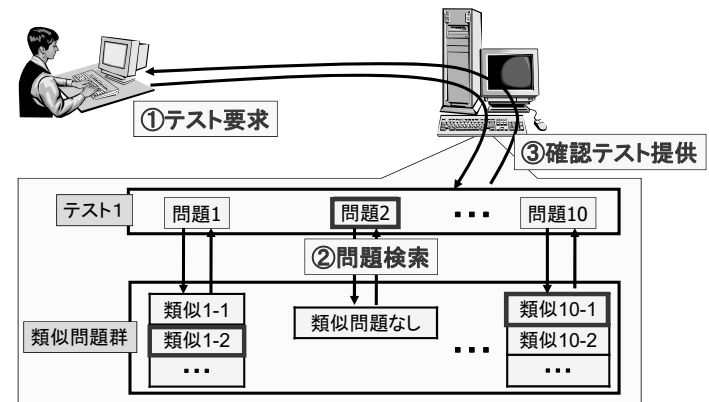


図 5 類似問題群からの適応的なテスト出題
Figure 5 Adaptive test setting from the similar quiz group.

5. 実験・評価

5.1 実験概要

本研究で提案する算出式の妥当性を検証するために難易度の推定実験を行った。実験には4.1節で分析した問題を対象とし、IRTと、提案手法により問題の難易度を推定した。推定結果から、IRTと提案手法による難易度の差の絶対値を誤差として算出した。表 4 に提案手法における重みの値を示す。この値は表 3 で示した出題パターンごとの難易度の差を基に設定した。重み 1 分の変化量は 0.1442 であり、これは表 3 における最大値から最小値までの値の範囲を 5 段階に分割したものである。これらの重みと変化量を用い、実際に難易度を推定し、その結果を比較・検討する。

表 4 出題パターンの組み合わせごとの重み

Table 4 Weight of each quiz types

出題パターン	Pa	Pb	Pc	Pd	Pe	Pf
Pa		0	-1	-3	1	0
Pb	0		0	-3	0	-2
Pc	1	0		-2	2	2
Pd	3	3	2		4	2
Pe	-1	0	-2	-4		-3
Pf	0	2	-2	-2	3	

5.2 実験結果

表 5 に出題パターンごとの実験結果、表 6 に比較問題数と誤差を示す。結果から、誤差の平均値は 0.359 となった。特に、Pd や Pe では誤差はほぼ無かった。また、比較数が 3 問以上の場合の誤差の平均値は 0.231 となり、比較問題数が多いほど誤差が少なくなった。このことから、一部の出題パターンや比較問題数が多い問題では、提案手法により IRT で推定した値と近い値が推定できることが分かった。

しかし、一部の問題では比較問題数が多い場合にも誤差が生じた。特に、比較的問題数の多い Pa や Pf では誤差の大きい問題が複数存在した。誤差の原因の一つとして、提案手法では選択肢の内容を考慮していないことが挙げられる。多肢選択式の問題では、明らかに誤りだと分かる選択肢や、紛らわしい選択肢などにより難易度が変化すると考えられるが、提案手法では問題の出題パターンのみに着目して難易度の推定を行うため、IRT と提案手法で推定した難易度に差が生じたと考えられる。

表 5 出題パターンごとの結果
Table 5 Results for each quiz types

出題パターン	比較個数平均	難易度平均		誤差平均	データ数
		本方式	IRT		
Pa	2.6	-0.858	-0.834	0.417	20
Pb	2.6	-1.010	-0.823	0.525	6
Pc	3.8	-0.757	-0.757	0.236	5
Pd	5.5	-0.456	-0.552	0.096	4
Pe	6.7	-0.869	-0.809	0.091	3
Pf	3.5	-0.885	-0.967	0.401	11
全体	3.4	-0.838	-0.833	0.359	49

表 6 比較問題数ごとの誤差
Table 6 Error for each quiz types

比較問題数	1	2	3	4	5	6	7	8	9	10
誤差平均	0.58	0.38	0.25	0.26	0.00	0.26	0.29	0.33	0.04	0.01

6. 選択肢の内容と難易度の分析

5 章の実験結果から、選択肢の内容を考慮した難易度推定法について検討する必要性が示唆された。多肢選択式の問題の場合、明らかに間違いと分かる選択肢や、紛らわしい選択肢などにより問題の難易度が変化すると考えられる。明らかに間違いと分かる選択肢が存在する場合、その選択肢を選択する学習者は少なくなり、また、すべての選択肢が紛らわしい場合、学習者の解答はばらけると考えられる。そのため、それぞれの選択肢が学習者に選択される確率（以下、選択率）と問題の難易度には関係性があると考えられる。

そこで、本研究では、講義「コンピュータネットワーク論」において 72 名の学生を対象にテストを行い、選択率と問題の難易度の関係性の分析を行った。分析には、CollabTest で過去に作成された問題 10 問を出題した。さらに、選択率の標準偏差を選択率のばらつきとし、問題の正答率と標準偏差の相関を調べた。

表 7 に問題ごとの各選択肢における選択率を示し、表 8 に各問題の正答率と標準偏差を示す。また、図 6 に正答率と標準偏差の分析結果を示す。正答率と標準偏差の相関係数は約 0.9735 となり強い相関が見られた。選択率の標準偏差が高い場合、特定の選択肢の選択率が高くなり、逆に標準偏差が低い場合は、すべての選択肢の選択率がほぼ均等になる。このことから、各選択肢の選択率と問題の難易度には関係があることが分かった。

表 7 各選択肢における選択率

Table 7 Selection probability of each choices

問題	問 1	問 2	問 3	問 4	問 5	問 6	問 7	問 8	問 9	問 10
選択肢 1	43%	15%	21%	7%	67%	8%	8%	49%	17%	25%
選択肢 2	18%	15%	17%	14%	15%	8%	39%	21%	24%	13%
選択肢 3	17%	18%	32%	75%	7%	6%	32%	10%	53%	33%
選択肢 4	21%	50%	28%	4%	10%	78%	21%	21%	7%	26%

表 8 各問題の正答率と標準偏差

Table 8 Correct answer rate and standard deviation of each choices

問題	問 1	問 2	問 3	問 4	問 5	問 6	問 7	問 8	問 9	問 10
正答率	43%	50%	32%	75%	67%	78%	32%	49%	53%	26%
標準偏差	7.73	10.6	4.27	20.9	17.6	22	8.34	10.3	12.3	5.41

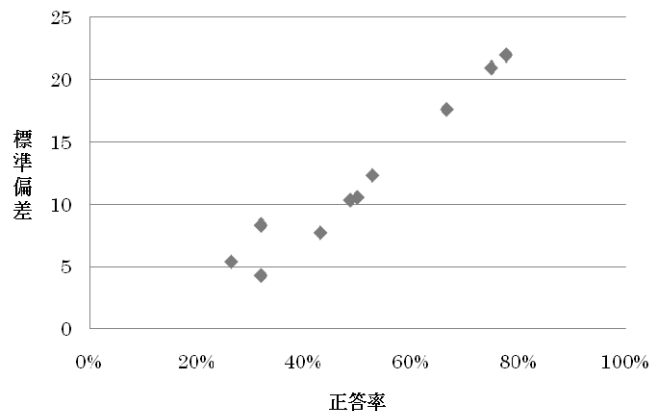


図 6 正答率と標準偏差の比較

Figure 6 Comparison of correct answer rate and standard deviation

7. 今後の課題

今後は、問題の選択肢の内容を考慮できるように提案手法を拡張することで、より精度の高い難易度の推定を目指す。特に、選択率と難易度に強い相関があることが分かったため、選択率に着目することで難易度推定の精度の向上が期待できる。しかし、選択率を分析するためには、予め問題を学習者に解答させなければならない。そのため、まずは出題パターンごとに選択肢を分析し、各出題パターンにおいて選択肢の内容が問題の難易度やそれぞれの選択肢の選択率にどのような影響を与えるのかを明らかにする。具体的には、問題文や選択肢内の専門用語に着目し、問題内で問われている知識とそれらの知識と関係のある専門用語の関連付けを行い、専門用語に基づく対象知識のクラスタリングや分類をする。そして、その分類結果と問題の難易度やそれぞれの選択肢の選択率との関係性を分析することで、問題の選択肢を考慮した難易度推定手法の検討を行う。

8. まとめ

本論文では、アイテムバンク内に存在する解答履歴のない問題の難易度を推定するために、問題の出題パターンに着目した難易度推定手法について検討し、実験・評価を行った。提案手法では、予め難易度が推定してある問題を複数用意することで、そ

れらの類似問題の難易度を出題パターンの組み合わせに基づき推定する手法である。出題パターンの変化に伴い、類似問題間で難易度にどのような差が生じるのかを明らかにするために、CollabTest で作成された問題を対象とし分析を行った。そして、その分析結果を基に独自の難易度算出式を用いた難易度推定手法を提案した。

IRT で推定した難易度との比較実験から、提案手法により、比較する問題数が多いほど IRT で推定した難易度と近い値が推定できることが分かった。しかし、一部には誤差の大きい問題も存在した。この誤差の原因の一つとして、提案手法では選択肢の内容を考慮していないことが挙げられる。多肢選択式の問題では、明らかに誤りだと分かる選択肢や、紛らわしい選択肢などにより難易度が変化すると考えられる。

今後は、それぞれの選択肢の選択率や問題の内の専門用語に着目し、出題パターンごとに選択肢の内容を考慮することのできる難易度推定手法の検討を行い、より推定の精度の高い手法を目指す。

参考文献

- [1] 植野真臣, 永岡慶三: e テスティング, 培風館, (2009).
- [2] 日本教育工学会編, 教育工学事典, 実教出版, (2000).
- [3] 豊田秀樹: 項目反応理論 入門編-テストと測定の科学, 朝倉書店, (2002).
- [4] Baker, F. B.: The Basics of Item Response Theory .2nd ed. USA, ERIC Clearinghouse on Assessment and Evaluation, (2001).
- [5] 高木輝彦, 高木正則, 勅使河原可海: 学生が作成した問題の出題パターンによる類似度算出手法の提案と評価, 情報処理学会教育シンポジウム(SSS2008)論文集, pp.95-102, (2008).
- [6] 高木輝彦, 高木正則, 勅使河原可海: 学生が作成した問題の類似度算出手法の提案と評価, 情報処理学会論文誌, Vol.50, No.10, pp.2426-2439 (2009).
- [7] M.TAKAGI and Y.TESHIGAWARA. : A WBT System Enabling to Create New or Similar Quizzes Collaboratively by Students, The Second IASTED International Conference on Education and Technology(ICET2006), Calgary(Canada), Proceedings of ICET2006,pp.263-268,2006.