

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

池田 信一^{1,a)} 高木 輝彦² 高木 正則³ 勅使河原 可海¹

受付日 2012年3月19日, 採録日 2012年10月10日

概要: 近年, テストを構成する問題を項目とし, 項目の難易度などのテスト実施後の統計データを用いてテスト構成を支援する研究が多数行われている. 項目の難易度の定量化には, 一般的に項目反応理論 (IRT) が用いられている. IRT で項目の難易度を推定するためには, あらかじめ多くの被験者に項目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項目を解答させることは困難であるため, 解答データのない項目の難易度の推定が課題となる. そこで, 本研究では, 解答データのない項目の難易度の推定を目的とし, 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を提案する. 本手法では, IRT を用いて推定した項目の難易度を基準にし, その類似項目の難易度を出題パターンの差異や選択肢の類似性から推定する. しかしながら, 出題パターンの差異や選択肢の類似性により, 難易度がどの程度変化するかは明らかになっていない. そのため, まず, 出題パターンと項目の難易度の関係性を分析した. また, 選択肢の類似性は選択肢の選択率と関連があることが分析の結果明らかになった. そこで, 各選択肢の選択率の推定手法を提案し, 推定の結果得られた選択率と項目の難易度の関係性を分析した. そして, その分析結果を基に難易度推定手法を提案する. 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が推定できる可能性を示した.

キーワード: 項目反応理論, e テスティング, 難易度, 類似項目

A Proposal and Evaluation of a Method of Estimating the Difficulty of Items Focused on Item Types and Similarity of Choices

SHINICHI IKEDA^{1,a)} TERUHIKO TAKAGI² MASANORI TAKAGI³ YOSHIMI TESHIGAWARA¹

Received: March 19, 2012, Accepted: October 10, 2012

Abstract: In recent years, a number of researches on the support of the composition of the test by using statistical data such as based on the difficulty level of items, which are the quizzes to constitute tests, are done. In general, Item Response Theory (IRT) is used to quantify the difficulty level of items. In order to estimate the difficulty level of the items, the items need to be answered by many learners in advance. However, it is hard to make them answer all quizzes in the item bank. Therefore, we proposed a method of estimating the difficulty level of un answered items. In our method, the value is estimated from the difference between item types and the similarity between choices. At first, IRT is used to estimate the difficulty level of items. Then the difficulty level of these similar items are estimated. However, it is not clarified the changes of the difficulty level depending on the item types and similarity of choices. Thus, firstly the relationship between item types and difficulty level of items are analyzed. Then the relationship between the similarity of choices and difficulty level of items are also analyzed. Next, from the result of these analyses, a method for estimating the difficulty level of items focused on item types and similarity of choices is proposed. The experiment results of comparing the proposed method with an IRT method showed that the value of the difficulty level of items using the proposed method is close to the value estimated using IRT.

Keywords: item response theory, e-testing, difficulty level, similar item

¹ 創価大学大学院工学研究科
Graduate School of Engineering, Soka University, Hachioji,
Tokyo 192-8577, Japan
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, Chofu, Tokyo 182-8585, Japan
³ 岩手県立大学ソフトウェア情報学部
Faculty of Software and Information Science, Iwate Prefec-
tural University, Iwate 020-0193, Japan
a) e11m5202@soka.ac.jp

1. 研究の背景と目的

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, 細分化された教育目標や各目標に対応した項目, また, 項目の正答率や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, テストの

品質を保証し、より信頼性の高いテストを実施することが可能となる。また、これらの統計データを用いて、テスト構成を支援する研究が多数行われている [3], [4], [5]。これらの研究では項目の難易度の定量化に項目反応理論 (IRT: Item Response Theory) [6] が用いられる。IRT で項目の難易度を推定するには、あらかじめ多くの被験者に項目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項目が存在するため、それらすべての項目に対して解答データを収集することは困難である。また、同じ項目を複数回テストに出題することで、項目の内容が被験者に漏えいしてしまう恐れがあるため、定期的に新規項目を追加し、アイテム・バンク内の項目を入れ替えることが望ましい。そのため、アイテム・バンク内に存在する解答データのない項目の難易度の推定が可能となれば、容易に新規項目をアイテム・バンクに追加することができ、コストや時間をかけずに膨大な量の項目の難易度を推定することができると考えられる。

以上から、本研究の目的は解答データのない項目の難易度の推定とする。本研究では、項目の難易度は項目内で出現する知識の問われ方 (以下、出題パターン) と、選択肢間の類似性によって変化する [7] と考え、難易度が既知の項目 (以下、基準項目) と同一の知識*1 を問う項目 (以下、類似項目*2) の難易度を、出題パターンの違いや選択肢の類似性の違いから生じる難易度の差に着目して推定する。なお、本研究では専門用語についての知識を問う多肢選択式の項目を対象とし、計算問題や図表を使用した項目は対象外とする。また、項目間の類似度算出手法 [8] を用いて項目間の類似度が算出されていることを前提とする。

以下、2 章では、IRT の概要とパラメータの推定における課題について述べ、3 章では、それらの課題を解決するための関連研究について述べる。4 章では、提案手法を実現するうえでの課題と、課題解決のためのアプローチについて述べる。5 章では、出題パターンに着目した難易度の推定、6 章では、選択肢の類似性に着目した難易度の推定について検討を行う。そして、検討結果を基に 7 章では本研究で提案する難易度推定手法について述べる。さらに、8 章では、IRT と提案手法で推定した難易度を比較し、提案手法の有効性を検証する。9 章で今後の課題について述べ、10 章で全体をまとめる。

2. IRT の概要と課題

近年、e テスティングの実用化にともない、テスト実施後の統計データを用いて、テストの得点予測や所要時間の予測などのテスト構成を支援する研究や、被験者の理解

度に適した難易度の項目を出題する適応型テスト (CAT: Computer Adaptive Testing) などに関する研究がさかんに行われている [5], [9], [10]。これらの研究において、項目の難易度の定量化には一般的に項目反応理論 (IRT: Item Response Theory) が用いられる。IRT は IRT モデルと呼ばれる統計モデルを用いて、各項目の統計的な性質を明らかにする。項目の性質は、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 (ICC: Item Characteristic Curve) で表される。被験者の能力特性を表す能力母数が θ である被験者 i の項目 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、ここでは本研究で対象とする 1 パラメータ・ロジスティックモデル (1PLM) について述べる。

$$P_j(\theta_i | b_j) = \frac{1}{1 + \exp\{-D(\theta_i - b_j)\}} \quad (1)$$

ここで、 D は $D = 1.7$ となる定数である。 b は難易度母数であり、難易度の値が大きいほど難しい項目となる。難易度の値は数学的には $-\infty$ 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 [6]。

図 1 に $P(\theta | 0)$, $P(\theta | 2)$, $P(\theta | -2)$ とした場合の ICC を示す。ここで、 $\theta = b$ に着目すると、正答確率 $p = 0.5$ となる。このことから、難易度はその項目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。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。

このモデルから、被験者のテストへの解答を基に被験者の能力母数 θ 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]。これらの母数をより正確に推定するためには、多くの被験者に項

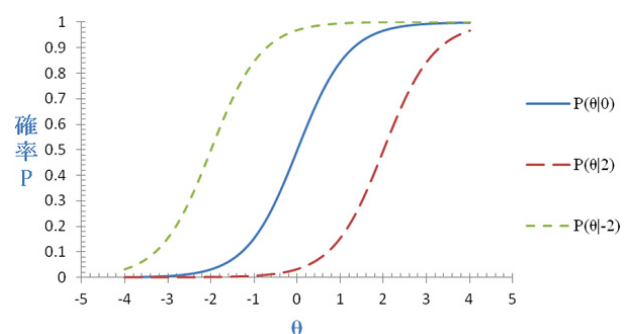


図 1 1PLM の項目特性曲線

Fig. 1 Item characteristic curve of 1PLM.

*1 分野特有の概念や法則、人物、歴史など。専門用語となることが多い (以下、対象知識)。

*2 本研究では類似項目を「項目内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項目」と定義する。

目を解答させる必要がある。一般的に、難易度に加え識別度や当て推量を考慮した上位モデルである 2PLM, 3PLM で正確な項目パラメータの推定を行うために必要な人数は、500 名から 1,000 名といわれている [11]。それに対し、1PLM では項目母数の数が少ないため、より少ない人数でパラメータの推定が可能である [12]。しかし、アイテム・バンク内に膨大な量の項目が含まれている場合、すべての項目に対する解答データを集めるのは困難であり、パラメータの推定に多大なコストと時間がかかってしまう。また、同じ項目を複数回テストに出題することで、項目内容が被験者に漏えいしてしまう恐れがある。そのため、項目は 1 回のみ出題され、定期的な新規項目をアイテム・バンクに追加することが望ましい。しかし、一般的に初回の出題はパラメータの推定のみ利用され、採点には利用されない。そのため、アイテム・バンク内に存在する解答データの無い項目の難易度の推定が課題となる。

3. 項目難易度推定の研究動向

2 章で述べた IRT における難易度推定の課題を解決するために様々な研究が行われている。豊田は既知となった項目の母数から未使用項目の母数を推定する手法を提案している [13]。この研究では、松原ら [14] による調査研究において検討されている、“責任”に関する 4 つの要因、“結果”、“関与”、“前歴”、“故意”を基に、‘母親’、‘安彦’、‘班長’、‘セールスマン’と題した 4 種類の刺激文を使用して項目母数を推定する実験を行い、モデルの有効性を示している。しかし、多肢選択式のようなテスト項目では、多くの要因が項目のパラメータと関連しているため、多肢選択式のような複雑な構造を持った項目には適用できない。

また、項目の形式や知識構造、選択肢の内容などに着目し、難易度を推定する手法や項目の自動生成についての研究が行われている [7], [10], [15]。桑原らは、プログラミング言語用の教材において、例題と演習問題の記述形式の変形度に着目し難易度を推定している [15]。この手法では、例題や演習問題に用いられたソースコードの知識構造を、言語仕様に基づいた各行の機能を構成する要素の記述形式（関数呼び出しや引数あり、引数なしなど）のツリーとして表現している。そして、例題の記述形式と演習問題の記述形式の変形度をツリー上での移動回数と定義し、その変形度の大きさから項目の難易度を推定するモデルを提案している。しかし、多肢選択式の項目の場合、このような知識構造を明確に定義するのは困難であると考えられる。菅沼らは、問題出題用のタグ（問題文や出現箇所、選択肢など）を埋め込んだ XML 文書を基に学生の理解度に応じた練習問題を自動生成するシステムを構築している [10]。このシステムでは、項目の出題形式を、選択問題、穴埋め問題、誤り訂正問題の 3 つに分類している。項目の難易度は、まず教師が難易度を見積もり、見積もった難易度をタグの属性

情報に設定する。そして、その値を基に 3 つの出題形式ごとに難易度の初期値を設定する。その後、学生への出題を繰り返すことで難易度を動的に変更していく。そのため、難易度の初期値は教師の主観による影響が強く、すべての項目の難易度を評価するのは教師の負担になってしまう。また、津森らは、語彙の親子関係を木構造で表現した概念空間を用いて語彙間の距離（概念距離）を算出し、語彙間の差異を考慮して選択肢を選び出すことで学習者の理解状況に適した難易度の項目を生成する手法を提案している [7]。この手法では、学習者が対象となる語彙を習得していない場合や、正解と不正解の語彙間の概念距離が小さいほど項目の難易度が高くなる。しかし、概念距離は 2 語間のリンクの本数のみで定義されているため、リンクの本数が同じ語彙間でも実際に学習者が感じる語彙間の差が異なる場合がある。また、同じ語彙を問う項目であっても、異なる問い方をすることで難易度が変化することも考えられる。

以上のように、項目の難易度推定手法に関する研究が多く行われているが、本研究のような専門用語に関する多肢選択式の項目の難易度を推定できる手法は著者らの知る限り存在しない。本研究は知識の間われ方に着目して項目を出題パターンに細かく分類したうえで、自然言語処理を用いて選択肢の類似性を算出することで難易度を推定する点に特徴がある。

4. 項目難易度に影響を及ぼす要因の考察

本研究では、IRT を用いて難易度を推定した項目を基準にし、それらの類似項目の難易度を推定する。我々はこの類似項目の難易度は基準項目との出題パターンの差異と選択肢の類似性によって変化すると考えた。以下に具体例を示して説明する。

4.1 出題パターンと項目難易度

我々は先行研究において、出題パターンに基づき項目を 11 の出題パターンに分類した [16]。項目の出題パターンは、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、出題パターンの詳細とそれらの例を示す。「その他」は、計算問題やある事柄の順序問題、穴埋め問題、または、不備のある項目などである。また、先行研究において、項目の管理方法として項目の類似性に着目し、項目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 [8]。この手法では、項目内で問われている知識の出現箇所を特定することで知識の抽出精度を向上させ、余弦により項目間の類似度を定量的な数値として算出している。表 2、表 3 に、異なる出題パターンの項目例を示す。

項目 1 ではスループットの説明を問題文で述べ、選択肢の中からスループットを選択させる項目であり、項目 2 で

表 1 項目の出題パターン

Table 1 Item patterns.

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項目 例：モールス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項目 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項目 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項目 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項目 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項目 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項目 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項目 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。
Pc+	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項目 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。
Pc-	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項目 例：以下に示す 4 つのネットワーク形態とその説明文のうち、誤っているものはどれか。
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項目 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）

表 2 異なる出題パターンの項目例 1（表 1 の出題パターン Pf）

Table 2 Example1 for the item with a different item pattern.

一定の時間内にシステムによって処理される仕事量を表す用語はどれか。
1. アクセスタイム
2. オーバーヘッド
3. スループット
4. ターンアラウンドタイム

表 3 異なる出題パターンの項目例 2（表 1 の出題パターン Pc+）

Table 3 Example2 for the item with a different item pattern.

スループットに関する記述のうち、適切なものはどれか。
1. ジョブとジョブの実行の間にオペレータが介入することによってシステムに遊休時間が生じて、スループットには影響を及ぼさない。
2. スループットは CPU 性能の指標であり、入出力の速度、オーバーヘッド時間などによって影響を受けない。
3. 多重プログラミングはターンアラウンドタイムの短縮に貢献するが、スループットの向上にはあまり役立たない。
4. プリンタへの出力を一時的に磁気ディスク装置へ保存するスプーリングは、スループットの向上に役立つ。

は、スループットに関する事柄について適した選択肢を選ばせる項目である。項目 1 の場合、スループットの意味を理解していれば解答することができるのに対し、項目 2 では、あらかじめスループットに関する知識を習得し、その知識を応用して解く必要がある。そのため、項目 1 よりも項目 2 の方が難易度が高くなると考えられる。

そこで、本研究では、あらかじめこれらの出題パターンに項目を分類し、類似項目間の出題パターンの変化にともなう難易度の差（以下、出題パターン成分の変化量）を定式化する手法について検討する。式 (3) に、2 つの異なる出題パターンの項目 i, j における、出題パターンに着目した難易度の変化量算出式を示す。

$$dp_{ij} = (D_{\max} - D_{\min}) \cdot sw_{ij} \quad (3)$$

ここで、 dp_{ij} は出題パターン成分の難易度の変化量を表す。 $D_{\max}$ はアイテム・バンク内の難易度の最大値であり、 $D_{\min}$ は難易度の最小値である。また、 sw_{ij} は難易度の値域 ($D_{\max} - D_{\min}$) に対する変化の割合である。つまり、出題パターンを比較する 2 つの項目のうち、一方の難易度がすでに分かっている場合、その難易度に式 (3) により算出した出題パターン成分の変化量 dp_{ij} を加えることで、もう一方の項目の出題パターン成分の難易度を推定することが可能となる。

しかし、式 (3) では sw_{ij} の値が既知でないと dp_{ij} は算出できないため、類似項目間における出題パターンの変化にともなう難易度の差を明らかにすることが課題となる。

4.2 選択肢間の類似性と項目難易度

多肢選択式の項目の場合、明らかに間違いと分かる選択肢が設定されている項目と、正答選択肢と概念や意味が類似・関連した誤答選択肢が設定されている項目では、難易度が変化すると考えられる。たとえば、選択肢どうしが概念上近いものである場合、被験者は誤って誤答選択肢を選択しやすくなると考えられる。逆に、正答選択肢と概念上異なる選択肢が含まれる場合、その選択肢を選ぶ被験者は少なくなると考えられる。したがって、正答選択肢に対して類似性の高い誤答選択肢ほど、学習者に選択される確率（以下、選択率）は高くなり、類似性が低い誤答選択肢ほど選択率は低くなると考えられる。なお、本研究における選択肢の類似性とは、選択肢が概念上どの程度似ているかを表し、単なる文字列としての類似性は考慮しないものとする。

以上より、本研究では、難易度が既知である項目とその類似項目間の選択率の差異によって、難易度の変化量（以下、選択肢成分の変化量）を算出する手法を検討する。式(4)に2つの類似項目 i, j における、各選択肢の選択率に着目した難易度の変化量の算出式を示す。

$$dc_{ij} = (D_{\max} - D_{\min})(cv_j - cv_i)dp_{\max} \quad (4)$$

ここで、 dc_{ij} は項目 j に対する項目 i の選択肢成分の難易度の変化量を表す。 $D_{\max}$, $D_{\min}$ は式(3)と同様アイテム・バンク内の難易度の最大値と最小値である。 v_i , v_j はそれぞれ項目 i, j における選択率の不偏分散の値であり、 c は選択肢の個数である。難易度の高い項目の場合、複数の誤答選択肢が選択される確率が高くなり、正答選択肢の選択率が低くなるため、選択率の分散値は小さくなると考えた。逆に、難易度の低い項目の場合、正答選択肢の選択率が高くなり誤答選択肢の選択率が低くなるため選択率の分散値は大きくなると考えられる。そのため、類似項目間で選択率の分散値を比較することで難易度の変化量を算出する。なお、 c 個の選択率 $x_1, x_2, \dots, x_c$ からなる不偏分散 v は、以下の式で算出することができる。

$$v = \frac{1}{c-1} \sum_{i=1}^c (\bar{x} - x_i)^2 \quad (5)$$

選択率の場合、各選択肢の選択率の合計が1であるため、式(5)により算出された分散値 v に選択肢の個数 c を掛けた値 cv の値域は、 $0 \leq cv \leq 1$ となる。また、 $dp_{\max}$ は式(3)の出題パターン成分の難易度変化量の最大値である。つまり、項目 i, j の分散値の差 $(v_i - v_j)$ の最大値は1となるため、分散値の差が最大るとき選択肢成分の難易度の変化量 dc_{ij} は $dp_{\max}$ と等しくなり、分散値の差が0のときは難易度の変化量も0となる。したがって、選択率を比較する2つの項目のうち、一方の難易度がすでに分かっている場合、その難易度に式(4)により算出した難易度の変

化量 dc_{ij} を加えることで、もう一方の項目の選択肢成分の難易度を推定することが可能となる。

しかし、あらかじめ被験者からの解答データが存在しなければ、選択率を算出することはできない。そのため、選択肢間の類似性に着目して各選択肢の選択率を推定する手法を検討する。

5. 出題パターンごとの項目難易度の傾向分析

5.1 分析の概要

本研究では、類似項目間における出題パターンの変化にともなう難易度の差を明らかにするために、IRTを用いて項目の難易度を推定したあと、出題パターンの異なる類似項目間における難易度の差を分析する。そして、その結果を基に式(3)の sw_{ij} を検討する。分析の手順を図2に示す。まず、IRTを用いて複数の項目（異なる出題パターンの類似項目）の難易度を推定する（図2中①）。次に、難易度を推定した項目間の類似度を算出し、それぞれの項目の類似項目を検索する（図2中②）。そして、出題パターンの異なる類似項目間の難易度の変化量を算出する（図2中③）。

難易度の差の比較では、誤答選択肢の差異による難易度の差が生じると考えられる。しかし、誤答選択肢は出題パターンによって形式が異なるため、すべての類似項目の誤答を等しくし、出題パターンの差異のみによる類似項目間の難易度変化量を抽出することが困難であった。そのため、本論文では出題パターンと選択肢が異なる類似項目間の難易度の差のみで分析を行った。なお、今後、選択肢成分の難易度の差を除外し、より正確な出題パターン成分の変化量を分析することを検討している。また、類似項目が存在しない項目は分析の対象外とし、出題パターンの異なる類似項目が存在する項目についてのみ分析を行った。分析で対象とする項目は、創価大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、2006年度から2008

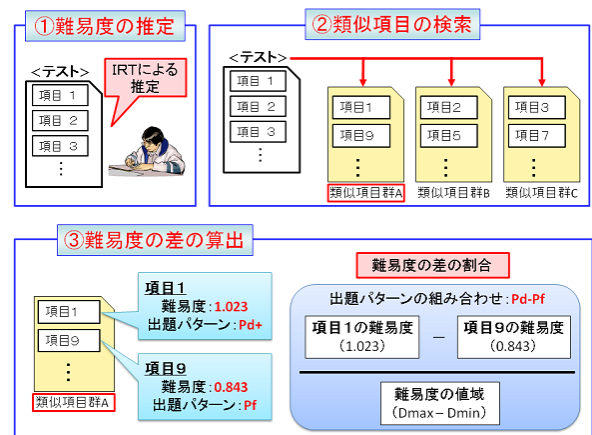


図2 出題パターンごとの難易度比較手順

Fig. 2 Difficulty level comparison procedure of each item pattern.

年度までに出题された 49 項目とした。これらの項目には各年度約 80 名の解答履歴データがあった。なお、表 1 の出题パターン ID に付与されている (+) と (-) の違いは、選択肢が正しいものか、誤っているものかだけであるため、この違いによる難易度の変化は生じないと考えられる。そのため、ここでは Pa~Pf の 6 つの出题パターンについて難易度の比較を行う。

5.2 難易度の変化量の算出

本研究では、項目 i 、項目 j 間の難易度の変化量 p_{ij} を以下の式で算出する。

$$p_{ij} = \frac{d_i - d_j}{D_{\max} - D_{\min}} \quad (6)$$

ここで、 d_i 、 d_j はそれぞれ項目 i 、項目 j の難易度を示す。 $D_{\max}$ は手順 ① で推定した難易度の最大値であり、 $D_{\min}$ は最小値である。つまり、 p_{ij} は手順 ① で推定した難易度の値域に対する、項目 i 、 j 間の難易度の差の割合を表す。そのため、 p_{ij} の値には以下の関係が成り立つ。

$$p_{ij} = -p_{ji} \quad (7)$$

そして、項目 i 、 j 間の出题パターンの組合せごとに、 p_{ij} の平均値を算出することで、出题パターンの差異による難易度の変化を分析する。

5.3 分析結果

表 4 に出题パターンの組合せごとの分析結果を示す。表のそれぞれの値は、行を項目 i の出题パターン、列を項目 j の出题パターンとしたときの p_{ij} の値である。なお、今回の結果では Pb と Pe の組合せは存在しなかった。分析結果から、類似項目間において出题パターンが変化することで、難易度に差が出る事が分かる。Pd や Pc の行に着目すると、難易度の変化量のほとんどが正の値を示しており、他の出题パターンに比べ難易度が高めに推定されている。一方で、Pa や Pe は難易度の変化量が負の値を示しており、他の出题パターンに比べて比較的難易度が低く推定された。これは、Pd や Pc は専門用語に関する事柄を問う項目であり、項目に解答する際に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。それに対し、Pa や Pf は、解答する際に知識を応用する必要はなく、選択肢が項目で問われている知識をそのまま表現している。また、Pe は各選択肢に記述された専門用語とその説明文の組合せから解答する項目であるため、各専門用語についての知識があれば解答することができる。そのため、これらの出题パターンは比較的難易度が低く推定されたと考えられる。以上の分析結果より、式 (3) では、表 4 の値を sw_{ij} の値として使用する。

6. 各選択肢の選択率推定手法の検討

本章では、各選択肢の選択率を推定する手法について検討を行う。多肢選択式項目の場合、「問題文で説明している内容と一致する選択肢はどれか」などのように、問題文の内容と選択肢の内容を正しく結び付けることで解答することのできる項目が一般的である。そのため、各選択肢が問題文に対して、どの程度類似しているのかによって選択率が変化すると考えられる。そこで、本研究では、問題文と選択肢の内容をテキスト情報としてベクトル空間で表現し、問題文に対する各選択肢の類似度を算出することで、各選択肢の選択率を推定する手法を検討する。

多肢選択式の項目では選択肢が専門用語のみである場合など、問題文や選択肢に含まれる語が少ない可能性がある。そのため、問題文と選択肢のテキスト情報のみでは類似度を算出することは困難である。そこで、本研究では、アイテム・バンクに蓄積されている項目から問題文や選択肢に関連する複数の専門用語（以下、関連語）を抽出する。項目内には対象知識となる専門用語とそれに関連する専門用語が含まれているため、アイテム・バンクに蓄積されている他の項目を用いることで、ある専門用語に対する関連語の抽出が可能であると考えられる。そして、抽出した関連語を含めたテキスト情報を基に問題文に対する各選択肢の類似度を算出する。

6.1 関連語の抽出手法

関連語の抽出の流れを図 3 に示す。まず、各項目から関連語の抽出を行う専門用語（以下、検索語）を決定する

表 4 出题パターンの組合せごとの難易度の差の平均値

Table 4 Averages of differences in the difficulty level by each combination of item patterns.

$i \backslash j$	Pa	Pb	Pc	Pd	Pe	Pf
Pa		0.011	-0.067	-0.219	0.116	-0.009
Pb	-0.011		0.010	-0.249		-0.155
Pc	0.067	-0.010		-0.179	0.122	0.145
Pd	0.219	0.249	0.179		0.336	0.136
Pe	-0.116		-0.122	-0.336		-0.251
Pf	0.009	0.155	-0.145	-0.136	0.251	

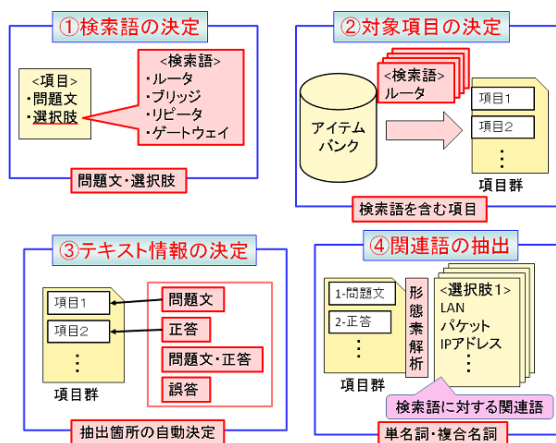


図 3 関連語の抽出手順

Fig. 3 Extraction procedure of related terms.

(図 3 中 ①). そして, その検索語を含むアイテム・バンク内の項目 (以下, 関連項目) を検索し (図 3 中 ②), 関連語の抽出箇所 (問題文・正答・誤答) を決定する (図 3 中 ③). 最後に, 決定された抽出箇所から関連語を抽出する. なお, 関連語の抽出には中川らによって開発された「専門用語抽出システム」[17] を用いる.

多肢選択式形式の場合, 検索語とする専門用語の出現箇所は選択肢の形式によって変わる. また, 関連項目内に検索語と直接関連しない選択肢が存在する場合があります. これらの選択肢の内容を関連語として使用してしまうと, 類似度算出の際のノイズとなってしまう. そこで, 本研究では選択肢を構成する要素を 2 つの形式に分類することで, その形式ごとに検索語の出現箇所を自動決定する. また, 関連項目内の検索語の出現箇所により, あらかじめ関連語の抽出箇所を絞り込むことでより正確に関連語を抽出する.

まず, 選択肢の形式と, その形式ごとの検索語の出現箇所の自動決定について説明する. 本研究では, 選択肢を構成する要素を以下の 2 つの形式に分類する.

- Type-T: 専門用語 “Term” (表 1 中 Pb, Pd, Pf)
- Type-S: 文章 “Sentence” (表 1 中 Pa, Pc, Pe)

表 2 の項目は Type-T となり, 表 3 の項目は Type-S となる. 以下に, 各形式の検索語の出現箇所と, 類似度算出の際に使用するテキスト情報を示す.

• Type-T

各選択肢の専門用語を検索語とする. したがって, 1 つの項目に対し選択肢の個数分の関連語集合が抽出される. そこで, 各選択肢に対する関連語集合と問題文に含まれる専門用語をテキスト情報として, 問題文に対する各選択肢の類似度を算出する.

• Type-S

対象知識を検索語として関連語を抽出する. そのため, 対象知識に対する関連語集合と各選択肢に含まれる専門用語の集合をテキスト情報として, 問題文に対する各選

表 5 検索語の出現箇所に基づく関連語の抽出箇所

Table 5 Extraction parts of the related terms based on the occurrence part of a search word.

出現箇所		抽出箇所
選択肢	正答	問題文と正答
	誤答	出現した選択肢
問題文		問題文と正答

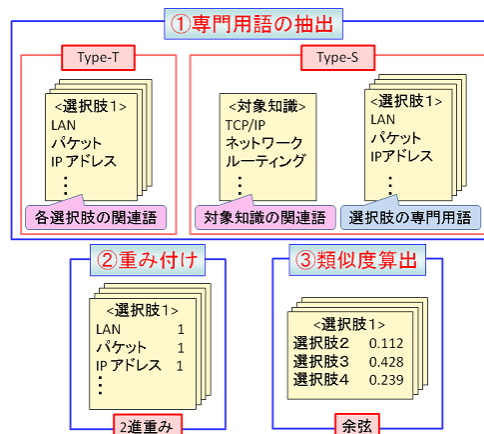


図 4 選択率の算出手順

Fig. 4 Calculation procedure of the selection probabilities.

択肢の類似度を算出する.

関連項目内の関連語の抽出箇所は正しいものを選ぶ項目と誤っているものを選ぶ項目で抽出箇所を変更する. 表 5 に関連項目が正しいものを選ぶ項目 (表 1 の出題パターンに + が付与されている項目) の場合, 関連語の抽出箇所を示す. 誤答選択肢の内容は問題文とは関係していないため, 誤答選択肢は関連語の抽出対象外とする. また, 関連項目が誤っているものを選ぶ項目 (表 1 の出題パターンに - が付与されている項目) の場合は, 正答選択肢が問題文で問われている内容と「誤っているもの」であるため, 表 5 の正答と誤答を入れ替えた箇所から抽出を行う. このように, 関連項目内の検索語の出現箇所により関連語の抽出箇所を絞り込むことで, より正確に関連語の抽出をする.

6.2 選択率推定手順

本節では, 6.1 節で述べた関連語の抽出手法を基に, 各選択肢の選択率を推定する手法について述べる. 図 4 に選択率の推定手順を示す. まず, 6.1 節の手順により, 各項目の検索語に対する関連語と比較対象となる語の抽出を行う (図 4 中 ①). そして, 抽出された語に対して重み付けを行う (図 4 中 ②). 現段階では, 最も単純な手法として 2 進重みを適用している. これは, 抽出されたすべての語に対して重み 1 を付与する. 最後に余弦 [18] により類似度を算出する (図 4 中 ③). 余弦による類似度 $\sigma(d_x, d_y)$ は, x_i, y_i をそれぞれ文書 d_x, d_y の索引語 i に対する重みあるいは索引語の存在を表す 2 値変数, T を索引語の総数と

した場合、次のような式で表される。

$$\sigma(d_x, d_y) = \frac{\sum_{i=1}^T x_i \cdot y_i}{\sqrt{\sum_{i=1}^T x_i^2 \times \sum_{i=1}^T y_i^2}} \quad (8)$$

なお、本手法では、問題文に対する各選択肢の類似度を算出するため、選択肢ごとに0~1の値で類似度が算出される。そのため、各選択肢の類似度の合計が1になるように正規化した値を、推定した選択率の値とする。式(4)では、この手法により推定した選択率を v_i, v_j とすることで、項目 i, j 間における選択肢成分の難易度変化量を算出することが可能となる。

7. 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手法

5章では2つの類似項目間における出題パターン成分の難易度変化量の算出方法を検討してきた。また、6章では、各選択肢の選択率を推定する手法により算出した選択率を基に、選択肢成分の難易度の変化量を算出する方法について検討した。本章では、5章、6章の検討結果を基に、本研究で提案する出題パターンと選択肢の類似性に着目した難易度推定手法について述べる。7.1節では、具体的な難易度推定手順と難易度の算出式について説明する。7.2節では、提案手法により期待される効果と、本手法の応用例について、適応型テストと作問支援システムの2つに応用した場合について述べる。

7.1 難易度推定手順

図5に本研究で提案する難易度算出手順を示す。まず、あらかじめ、アイテム・バンク内で解答データを十分に保持する項目の難易度を、IRTを用いて推定する(図5中①)。次に、解答データのない項目(以下、推定項目)の類

似項目を検索し、その中から手順①で難易度が推定された項目(基準項目)を抽出する(図5中②)。そして、推定項目と基準項目の出題パターンの組合せを基に、式(3)により出題パターン成分の難易度の変化量を求める(図5中③)。なお、式(3)における変化の割合 sw_{ij} は①で難易度を推定した結果を基に設定する。また、6章で検討した手法により、推定項目と基準項目それぞれの選択率の分散値を算出し、その分散値を基に式(4)により選択肢成分の難易度の変化量を求める(図5中④)。最後に、③と④で算出した難易度の変化量を、基準項目の難易度に足すことで推定項目の難易度を算出し、その結果をアイテム・バンクに登録する(図5中⑤)。式(9)に本研究で提案する難易度算出式を示す。

$$PD = \frac{1}{n} \sum_{j=1}^n (ds_j + dp_{ij} + dc_{ij}) \quad (9)$$

ここで、 PD は n 間の基準項目 j から算出した推定項目 i の難易度である。 dp_{ij} は項目 i, j における出題パターン成分の難易度の変化量を表し、式(3)により算出する。また、 dc_{ij} は同様に選択肢成分の難易度の変化量を表し、式(4)により算出する。 ds_j はIRTにより推定した基準項目 j の難易度である。

難易度推定の流れは、既知である基準項目 j の難易度 ds_j に出題パターン成分の変化量 dp_{ij} と選択肢成分の変化量 dc_{ij} を足すことで、基準項目 j に対する推定項目 i の難易度を算出する。同じ処理を残りの $n-1$ 間の基準項目に対して行う。最終的に算出された難易度の平均値を推定項目 i の難易度とする。

同様の処理を他の類似項目にも行う。提案手法により難易度を推定した項目も、他の類似項目の基準項目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、基準項目の数が増える。したがって、提案手法では繰り返し難易度の推定を行い、前回の推定値との推定誤差があらかじめ設定した閾値より小さくなったところで、その項目の推定を終了する。

7.2 期待される効果と応用例

本研究で提案する難易度推定手法では、あらかじめIRTで複数の項目の難易度を推定することで、それらの類似項目の難易度を被験者からの解答データを集め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・バンクに新規項目を追加する際も解答データを集める必要がないので、アイテム・バンク内の項目数を容易に増やすことができる。アイテム・バンク内に難易度が既知である項目が増えれば、より幅広いテスト構成が可能となる。また、解答データのない項目の難易度推定のほかにも、出題パターンや選択肢の類似性による難易度の変化に着目することで、様々な応用例が考えられる。

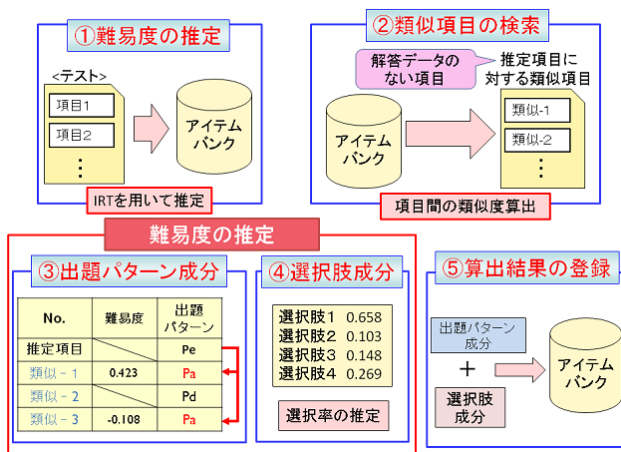


図5 難易度算出手順

Fig. 5 Calculation procedure of the difficulty levels.

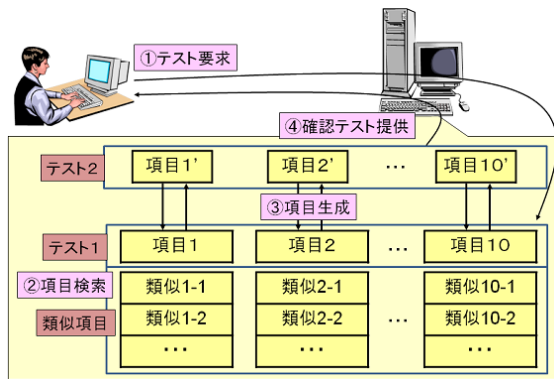


図 6 項目類似度データを用いた適応的なテスト出題

Fig. 6 Adaptive test setting from the similar item group.

まず 1 つは、オンラインテストへの応用が考えられる。近年、e テスティング技術において、適応型テスト (CAT: Computer Adaptive Testing) が注目されており、適応型テストを用いて学習者の学習意欲や学習効果の向上を目的とした問題演習システムに関する研究がなされている [19], [20]. 一方, Takagi らは学生による類似項目の作成が可能なシステムの開発を行い, 実験やアンケート結果からテストを複数回繰り返して解答する際に, 同一項目を出題するのではなく, 類似項目をランダムに出題することによる学習効果を示している [21]. しかし, この研究ではランダムに類似項目を出題するため, 学習者の理解状況に適した難易度の出題は行われていない。

そのため, 本研究で提案する類似項目間の出題パターンや選択肢の類似性による難易度の変化量に着目することで, より効果的なオンラインテストを実現することが可能であると考えられる。図 6 に項目類似度データを用いた適応的なテスト出題方式を示す。まず, 学生がテストを要求すると (図 6 中 ①), あらかじめ算出されている項目類似度データを基に各項目の類似項目を自動で検索する (図 6 中 ②)。そして, それらの項目の出題パターンや選択肢の内容を基に, 学生の理解度や解答状況に適した項目を生成し (図 6 中 ③), その項目を学生に提示する (図 6 中 ④)。この方式により, 毎回異なる出題パターンの類似項目を出題することができる。また, あらかじめ複数の選択肢候補を用意し, 選択率の異なる項目を生成することで, 学習者に適した難易度の項目を出題することが可能となると考えられる。そのため, 類似項目を繰り返し解くことによる知識の定着や, 様々な出題パターンの項目を解くことで, 同一知識に対して異なる角度からの学習が可能となり, より高い学習効果が期待される。

もう 1 つの利用例として, 作問時での利用が考えられる。多肢選択式項目の場合, 正答選択肢と類似した誤答選択肢を考え出す作業が作問者にとって大きな負担となる。そこで, 項目作成時に類似項目の選択肢を誤答選択肢の候補として提示することで作問を支援することができると考えら

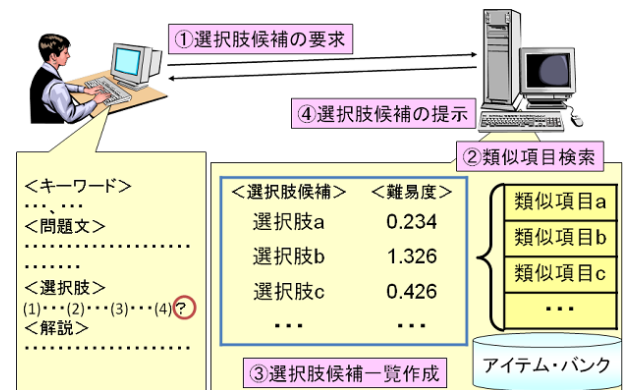


図 7 選択肢候補の要求と提示

Fig. 7 Requesting and showing the choice candidates.

れる。図 7 に選択肢候補の要求と提示の流れを示す。作問者が選択肢候補を要求すると (図 7 中 ①), すでに入力されている項目情報から類似項目を自動で検索する (図 7 中 ②)。そして, 類似項目の選択肢を抽出し, 選択肢候補の一覧を作成し (図 7 中 ③), 作問者へと提示する。その際, 各選択肢候補の選択率を推定し, 提示した候補による難易度の変化を表示することで, 作問者は自身が望む難易度の項目となるような選択肢を選ぶことが可能となると考えられる。たとえば, 作問者がより難易度の高い項目を作成したい場合, 推定された選択率の高い選択肢を提示し, 逆に, 難易度の低い項目を作成したい場合は, 選択率の低い選択肢を提示する。これにより, 誤答選択肢の作成が容易になり作問者の負担軽減が期待できる。

8. 評価

8.1 実験概要

本章では, 前章で提案した難易度推定手法の妥当性を検証するために難易度の推定実験を行う。実験では初級システムアドミニストレータ試験に'04 年度から'08 年度までに出题された項目, および, 基本情報技術者試験に'07 年度から'10 年度までに出题された 1,000 項目からなるアイテム・バンクをあらかじめ用意しておく。アイテム・バンク内の項目は, 問題文・正答選択肢・誤答選択肢・出題パターンを分けて管理してある。実験方法は, 5 章で述べた出題パターン成分の難易度の変化量のみを考慮した場合 (以下, D_p) と 6 章で述べた選択肢成分の難易度の変化量のみを考慮した場合 (以下, D_c), この 2 つを考慮した提案手法 (以下, $D_p + D_c$) の 3 つの手法と IRT それぞれの手法を用いて難易度を推定し, 3 つの手法で算出した難易度と IRT で推定した難易度の差と相関係数を算出する。提案手法を用いた難易度の推定では, 式 (3) により出題パターン成分の難易度の変化量 dp_{ij} を算出する際に必要となる, 難易度の変化の割合 sw_{ij} は表 4 の値を使用する。式 (4) を用いた選択肢成分の難易度の変化量 dc_{ij} の算出では, アイテム・バンクに含まれている 1,000 項目を基に関連語を抽出し, 各

選択肢の選択率を推定する。そして、推定した選択率の分散値を基に dc_{ij} を算出する。難易度推定の手順は、まず、アイテム・バンクに含まれている 20 項目から構成される事前テスト（以下、テスト 1）を実施し、IRT を用いて難易度を推定する（図 5 中 ①）。IRT を用いた難易度の推定では 1PLM を使用する。このテスト 1 の項目を提案手法における難易度推定の際の基準項目として、今回の実験では創価大学および岩手県立大学、電気通信大学の 3 大学の学生 82 名に対してテスト 1 を実施し、その解答データを使用して難易度の推定を行う。次に、テスト 1 に出題された項目の類似項目 20 項目を検索し（図 5 中 ②）、それらの類似項目からなるテスト（以下、テスト 2）を実施し、IRT を用いて難易度を推定する。今回は、テスト 1 とテスト 2 の結果を同一の尺度上の値に変換するため、共通項目 5 項目を使用した共通項目デザインにより等化 [12] を行う。そして、テスト 2 に含まれる項目の難易度を D_p , D_c , D_p+D_c の 3 つの手法で算出し、それぞれの手法で算出した難易度と IRT で推定した難易度の差と相関係数を算出する。

8.2 実験結果と考察

表 6 にそれぞれの手法により推定した難易度の一覧、表 7 に IRT で推定した難易度と 3 つの提案手法の差を示す。なお、表に示す結果は推定項目（テスト 2）の推定結果であり、等化で使用した共通項目 5 項目は除いてある。IRT で推定した難易度と、各手法で推定した難易度の相関係数は、 $D_p : 0.46$, $D_c : 0.12$, $D_p+D_c : 0.37$ となり、 D_p と D_p+D_c においてやや強い正の相関がみられたことから、提案手法で推定した難易度と IRT で推定した難易度には関係性があることが分かった。しかし、 D_c ではあまり相関はみられなかった。また、提案手法と IRT で推定した難易度の順序から、Spearman の順序相関係数を算出した結果、0.31 となり IRT で推定した難易度と、提案手法で算出した難易度の相関はあまり強くないと考えられる。

一方で、比較的大きかった項目 4、項目 5、項目 6、項目 9、項目 11 の 5 項目を除いた計 10 項目において、IRT で推定した難易度と各手法で推定した難易度の相関係数を

調べた結果、 $D_p : 0.72$, $D_c : 0.76$, $D_p+D_c : 0.80$, また、Spearman の順序相関係数は 0.67 となり、強い正の相関がみられた。そのため、差の大きかった項目以外では、提案手法を用いて IRT で推定した難易度と近い値を推定できる可能性があることが分かった。

差の大きい項目の原因としては、 D_c の相関が最も低いことから、選択率の推定結果に原因があると考えられる。そこで、差の大きかった 5 問の選択率に着目する。表 8 に差の大きい項目における解答選択率と提案手法で推定した選択率を示す。なお、ここではテスト 1、テスト 2 の結果において選択率の推定値が実際にテストで測定した選択率と異なるテストの方のみを示す。テスト 1 の項目 4、項目 5、テスト 2 の項目 6 は正答率が 3 割以下の難しい項目であった。特に項目 4、項目 6 では解答選択率において誤答の選択肢が最も高くなった。提案手法では、推定した選択率の分散値を基に難易度の変化量を算出する。しかし、正答選択肢の選択率が 0.8 の項目と誤答選択肢の選択率が 0.8 となるような項目の場合、2 つの項目の難易度は大きく異なるが、分散値は近い値となってしまう。そのため、誤答選択肢の選択率が最も高くなる項目では、選択率の分散値のみで難易度の変化量を適切に算出できず、差が大きくなったと考えられる。また、項目 5 では選択肢 4 の選択率が低く推定され、分散値が大きくなったため差が生じたと考えられる。テスト 2 の項目 9 は、正答率が約 9 割のやさしい項目であった。しかし、誤答選択肢である選択肢 4 の選択率が 0.33 と比較的高い値で推定されたため、結果として差が大きくなったと考えられる。一方、提案手法では関連語が存在せず選択率の推定値が 0 となった項目があった。テスト 2 の項目 6 では、誤答選択肢の関連語が存在せず、正答選択肢の選択率が 1 となってしまった。テスト 1 の項目 11 では正答選択肢の「ルータ」に対して、誤答選択肢の「ゲートウェイ」や「リピータ」は関連語が存在せず選択率の推定値が 0 となった。しかし、これらの選択肢は概念上近い意味の専門用語であり、実際の解答選択率はともに約 0.2 となったため結果として差が大きくなったと考えられる。そのため、関連語の存在しない選択肢に対する選択率

表 6 難易度推定結果

Table 6 Estimation results of the difficulty levels.

項目		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
難 易 度	Dp	0.50	0.78	-0.11	1.37	0.45	0.85	-1.95	-0.57	-1.30	0.76	0.64	0.34	-0.73	-1.50	0.27
	Dc	-0.25	0.80	-0.60	1.26	0.78	-0.52	-1.88	0.65	-0.72	0.65	0.86	0.26	-0.96	-1.60	0.30
	Dp+Dc	0.48	1.29	-0.11	1.26	0.18	0.22	-1.85	0.05	-0.94	0.65	0.83	0.49	-1.18	-1.60	0.27
	IRT	-0.66	1.40	-0.26	-1.90	-2.36	4.35	-1.90	-0.26	-4.10	0.96	-2.36	-1.47	-1.90	-1.90	-1.06

表 7 各手法ごとの IRT で推定した難易度との差

Table 7 The differences in the difficulty levels of each method.

項目		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
の 差 難 易 度	Dp	1.16	0.61	0.15	3.27	2.81	3.50	0.05	0.31	2.80	0.20	3.00	1.82	1.17	0.41	1.33
	Dc	0.41	0.60	0.33	3.16	3.14	4.87	0.02	0.91	3.39	0.31	3.22	1.73	0.95	0.30	1.36
	Dp+Dc	1.15	0.11	0.15	3.16	2.54	4.13	0.05	0.31	3.16	0.31	3.19	1.96	0.72	0.30	1.33

表 8 差の大きい項目の選択率

Table 8 Selection probabilities of items with large differences.

選択率 (観測値)		選択肢 1	選択肢 2	選択肢 3	選択肢 4
項目 4	テスト 1	0.27	0.43	0.17	0.13
項目 5	テスト 1	0.30	0.21	0.22	0.27
項目 6	テスト 2	0.21	0.13	0.57	0.09
項目 9	テスト 2	0.88	0.04	0.04	0.04
項目 11	テスト 1	0.21	0.17	0.24	0.38
選択率 (推定値)		選択肢 1	選択肢 2	選択肢 3	選択肢 4
項目 4	テスト 1	0.65	0.35	0	0
項目 5	テスト 1	0.39	0.25	0.31	0.05
項目 6	テスト 2	0	1	0	0
項目 9	テスト 2	0.50	0.17	0	0.33
項目 11	テスト 1	0	0.26	0	0.76

の推定について検討する必要がある。また、テスト 1 とテスト 2 における IRT で推定した難易度の差を算出し、その値と表 7 の差 ($D_p + D_c$) の相関係数を調べた結果、0.973 と強い相関がみられた。このことから、基準項目と推定項目における難易度の変化量が大きいほど、提案手法では差が大きくなることが分かった。

9. 今後の課題

8 章の実験結果から、提案手法において選択率が正しく推定されていない場合に、IRT で推定した難易度との差が大きくなることが分かった。選択率は、関連語が存在せず選択率が 0 となった場合や、誤答選択肢の選択率が最も高くなる場合などが原因で正しく推定されなかった。そのため、アイテム・バンク内に関連項目が存在しない専門用語の場合は、あらかじめいくつかのキーワードを設定し、それらを関連語とする必要があると考えられる。また、誤答選択肢の選択率が最も高くなるような項目の場合、選択率の分散値のみでは正確に難易度の変化量が算出できないことが分かった。そのため、より正確な難易度変化量の算出式を検討する必要がある。さらに、今回はすべての関連語の重みを等しく扱った。しかし、本研究で対象としている専門用語について問うような項目では、解答に必要となる知識とより密接に関わる専門用語が項目内に含まれている場合も考えられる。そのため、今後は関連語に対する適切な重み付け手法についても検討を行う。

また、今後は提案手法を用いてアイテム・バンク内の項目の難易度を推定するツールを開発し実装する。今回は初級システムアドミニストレータや基本情報処理技術者の資格試験を対象としたが、異なる分野においても、専門用語に関する選択問題で表 1 のような出題パターンであれば、提案手法を適用することは有効であると考えられる。

10. まとめ

本論文では、アイテム・バンク内に存在する解答データ

のない項目の難易度の推定を目的として、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手法について検討し、実験・評価を行った。提案手法では、あらかじめ難易度が推定してある項目を複数用意することで、それらの類似項目の難易度を出題パターンや選択率の分散の差異に着目して推定する。まず、出題パターン成分の難易度変化量の算出のため、出題パターンの変化にともない類似項目間で生じる難易度の差を分析した。また、選択肢成分の難易度変化量算出のため、問題文や選択肢に含まれる専門用語の関連語を用いて、解答データのない項目の選択率を推定する手法を提案した。そして、出題パターン成分の難易度変化量と選択肢成分の難易度変化量を基に、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、提案手法により出題パターンと選択肢の類似性に着目することで、IRT で推定した難易度と近い値が推定できる可能性を示した。そのため、アイテム・バンクに新規項目を追加する際に被験者からの解答データを集めずに難易度の推定が可能となり、アイテム・バンク構築の際にかかるコストや時間の軽減につながると期待される。しかし、一部の項目は選択率が正しく推定されず、差が大きくなってしまった。今後は、基準項目の難易度を考慮に入れた選択肢成分の難易度算出式や、適切な選択率推定手法について検討を行いより精度の高い難易度の推定手法を目指す。

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] ソンムアン・ボクボン, 植野真臣: e テスティングにおける得点・時間予測システムの開発, 電子情報通信学会論文誌, Vol.J91-D, No.9, pp.2225-2235 (2008).
- [4] 大友賢二: 項目応答理論—TOEFL・TOEIC 等の仕組み, 電子情報通信学会誌, Vol.92, No.12, pp.1008-1012 (2009).
- [5] 高橋 登, 中村知靖: 適応型言語能力検査 (ATLAN) の作成とその評価, 教育心理研究, Vol.57, No.2, pp.201-211

- (2009).
- [6] Baker, F.B.: *The Basics of Item Response Theory*, 2nd ed., ERIC Clearinghouse on Assessment and Evaluation, USA (2001).
 - [7] 津森伸一, 海尻賢二: 理解状況に適応した選択式問題生成方法の検討, 教育システム情報学会誌, Vol.26, No.3, pp.240-251 (2009).
 - [8] 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).
 - [9] 田中 浩, 糸賀裕弥, 毛利公一, 山本哲男, 島川博光: 個々の学習者の理解状況と学習意欲に合わせたプログラミング教育支援, 情報処理学会論文誌, Vol.48, No.2, pp.958-968 (2007).
 - [10] 菅沼 明, 峯 恒憲, 正代隆義: 学生の理解度と問題の難易度を動的に評価する練習問題自動生成システム, 情報処理学会論文誌, Vol.46, No.7, pp.1810-1818 (2005).
 - [11] Ayala, R.J.: *The Theory and Practice of Item Response Theory*, Guilford, New York (2009).
 - [12] 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).
 - [13] 豊田秀樹: 未使用のラッシュ型項目の母数の推定, 心理学研究, Vol.73, No.1, pp.26-33 (2002).
 - [14] 松原 望: 人間行動の計量分析, 第 10 章, 東京大学出版会 (1990).
 - [15] 桑原恒夫, 中村喜宏, 玉城幹介, 山田光一: コンピュータ言語用教材の演習問題の難易度と例題からの変形度との関係, 日本教育工学雑誌, Vol.23, No.3, pp.135-146 (1999).
 - [16] 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).
 - [17] 東京大学中川研究室, 横浜国立大学森研究室: 専門用語自動抽出システム.
 - [18] 徳永健伸: 情報検索と言語処理, 東京大学出版会 (1999).
 - [19] 加藤 浩, 赤堀侃司: ベイズ推定による適応的問題演習システムのための問題選択方式, 電子情報通信学会論文誌 D-II, Vol.J82-D-II, No.1, pp.147-157 (1999).
 - [20] 渡辺佳代, 岡田美保子, 原平八郎, 寺延美恵子: 電子カルテ学習支援システムにおけるテストの評価に関する研究, 川崎医療福祉学会誌, Vol.20, No.1, pp.213-221 (2010).
 - [21] Takagi, M. and Teshigawara, Y.: A WBT System Enabling to Create New or Similar Quizzes Collaboratively by Students, *The 2nd IASTED International Conference on Education and Technology (ICET2006)*, Proc. ICET2006, Calgary (Canada), pp.263-268 (2006).



池田 信一 (学生会員)

2011 年創価大学工学部情報システム工学科卒業。現在, 同大学大学院博士前期課程在学中。e-Learning, e テスティング, 情報検索, 自然言語処理等の研究に従事。情報処理学会第 73 回全国大会学生奨励賞受賞。人工知能学

会会員。



高木 輝彦 (学生会員)

2010 年創価大学大学院工学研究科情報システム工学専攻博士前期課程修了。現在, 電気通信大学大学院情報システム学研究科社会知能情報学専攻博士後期課程在学中。e-Learning, 教育工学, 情報検索, 自然言語処理, テキ

ストマイニング等の研究に従事。情報処理学会第 70 回全国大会学生奨励賞, マルチメディア, 分散, 協調とモバイル (DICOMO2008) シンポジウムヤングリサーチ賞, 日本テスト学会第 10 回大会大会発表賞受賞。人工知能学会, 日本教育工学会, 教育システム情報学会, 言語処理学会各会員。



高木 正則 (正会員)

2003 年創価大学工学部情報システム学科卒業。2007 年同大学大学院工学研究科博士後期課程修了。博士 (工学)。同大学工学部助教を経て現在, 岩手県立大学ソフトウェア情報学部

講師。e-Learning, e テスティング, e ポートフォリオ, CSCL 等の研究に従事。情報処理学会第 65 回全国大会学生奨励賞, 情報処理学会情報教育シンポジウム (SSS2003) 奨励賞受賞。教育システム情報学会, 日本教育工学会, 日本テスト学会, 環境情報科学センター, 日本福祉介護情報学会各会員。



勅使河原 可海 (フェロー)

1970 年東京工業大学大学院理工学研究科博士課程修了。工学博士。同年日本電気入社。コンピュータネットワーク, ネットワークアーキテクチャ, 衛星データネットワーク等の開発に従事。1974~1976 年ハワイ大学アロハ

システム客員研究員。技術企画部主席技師長, 技術戦略室長代理を歴任。1995 年創価大学工学部教授, 工学部長, 工学研究科長を歴任。ユビキタスコンピューティング, グループウェア, e-Learning, ネットワークセキュリティ等の研究に従事。著書に『コンピュータネットワーク』, 朝倉書店, 1983 年等。オペレーションズリサーチ学会フェロー, 電子情報通信学会, 経営情報学会, IEEE, ACM 各会員。