

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 提案

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

概要

問題の難易度の定量化には、一般的に項目反応理論 (IRT) が用いられている。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、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、解答データのない問題の難易度の推定が課題となる。そこで、本研究ではこれまで、解答データ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提案してきた。これは、項目反応理論 (IRT) を用いて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の組み合わせを基に推定する手法である。しかしながら、問題の難易度は選択肢の類似性によって変化することも考えられる。そこで、本論文では、選択肢の内容を考慮に入れた難易度推定手法を提案する。

A Proposal 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}

Abstract

In general, Item Response Theory (IRT) is used to quantify the difficulty level of quizzes. In order to estimate the difficulty level 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. Therefore, a method of estimating the difficulty level of unanswered quizzes by focusing on the quiz types is proposed. This method estimates the difficulty level based on the combination of the quiz types by using IRT to estimate difficulty level of some quizzes. However, the difficulty level of the quiz might change by the similarity of choices. Therefore, in this paper, a method of estimating the difficulty level in considering the contents of choices.

1. はじめに

近年、Web上で実施するテストである「eテスト」が注目されている[1]。eテストでは、細分化された教育目標や各目標に対応した問題、また、問題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]。これらの研究では問題の難易度の定量化に項目反応理論 (Item Response Theory: IRT) [6] が用いられる。IRT で問題の難易度を推定するには、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データのない問題の難易度の推定が課題となる。

この課題を解決するために、我々は、解答

^{†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 System, 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

データの無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した[7]。具体的なアイデアは、同一知識を問う問題（以下、類似問題）の出題パターンの違いから生じる難易度の差に着目し、難易度が既知の問題から別の出題パターンの問題の難易度を推定するというものである。この知識とは、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（以下、対象知識）。また、問題間の類似度を算出することで、類似問題を検索することが可能となっている[8]。

しかしながら、問題の難易度は出題パターンのみならず、選択肢の内容によって変化することが、津森らの研究によって報告されている[9]。例えば、問題文で問われている内容とは明らかに異なる選択肢が存在する場合や、どの選択肢も意味的に類似している場合などによって、問題全体の難易度が変化すると考えられる。

そこで、本研究では、選択肢の内容を考慮に入れた難易度推定手法を確立するために、選択肢と難易度の関係性を分析し、選択肢間の類似度算出手法について検討を行い、より精度の高い難易度の推定を目指す。なお、以降では、選択肢間の類似度とは正答選択肢に対する誤答選択肢の類似度とする。

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、本研究で提案する難易度推定手法の説明をし、4章では、提案手法における検討課題と、それに対するアプローチについて述べる。5章では、選択肢間の類似度算出手法について説明し、6章で類似度算出手法の妥当性の検証実験を行い、その結果を考察する。最後に、7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 $P_j(\theta)$ 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。

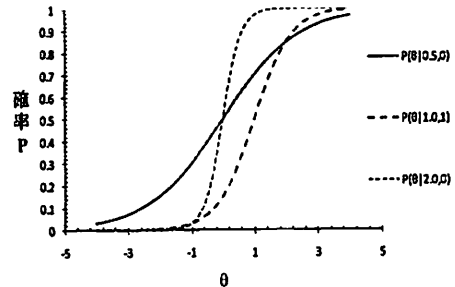


図1 2PLMの項目特性曲線

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

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、実質的に識別力は正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[6]。

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 個の問題に対する解答を $\mathbf{u}_i = \{u_{i1}, u_{i2}, \dots, u_{ij}, \dots, u_{in}\}$ 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 が問題 j に誤答する確率である。

このモデルから、学習者のテストへの解答

表 1 問題の出題パターン

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選び。
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）

を元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. 出題パターンに着目した難易度推定の検討

3.1 問題の出題パターン

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[10]。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。

3.2 出題パターンに基づく難易度推定手順

図 2 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・バンク内で解答データを十分に保持する問題の難易度を、

IRT を用いて推定する (図 2 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 2 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 2 中③)。さらに、①で推定した比較問題の難易度に変化量を足すことで、出題パターン成分の難易度を算出する (図 2 中④)。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。そして、選択肢間の類似度を基に選択肢成分の難易度を算出し (図 2 中④)、最終的に③と④で算出した難易度を足した値を推定問題の難易度としてアイテム・バンクに登録する (図 2 中⑤)。また、式 (3) に本研究で提案する難易度算出式を示す。

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

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_j 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。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。 dc は、選択肢成分の難易度であり、具体的な推定方法は今後検討していく。

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に選択肢成分の難易度 dc を足した値を推定問題 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3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。

4. 検討課題とアプローチ

本論文では、3.2節の難易度推定手法を確立するために正答選択肢に対する誤答選択肢の類似度算出手法について検討を行う。なお、選択肢を構成する要素は大きく分けて以下の2つの形式に分類できる。

- 形式F：専門用語（表1中Pb, Pd, Pf）
- 形式A：専門用語に関連する文章（表1中Pa, Pc, Pe）

以下に、各形式の問題の例を示す。

<問題1>

ネットワークの輻輳に対しても動的に対応するための機構としてトラフィックの流れを制御するのは次の制御のうちどれか。

- | |
|----------|
| 1. フロー制御 |
| 2. 順序制御 |
| 3. 誤り制御 |
| 4. 優先制御 |

<問題2>

トランスポート層を説明した文章を選べ。

- | |
|-------------------------------------|
| 1. ネットワークの端から端までの通信管理（エラー訂正、再送制御等）。 |
| 2. ネットワークにおける通信経路の選択（ルーティング）、データ中継。 |
| 3. 直接的に接続されている通信機器間の信号の受け渡し。 |
| 4. 通信プログラム間の通信の開始から終了までの手順。 |

問題1では、フロー制御の説明を問題文で述べ、選択肢の中からフロー制御を選択させる

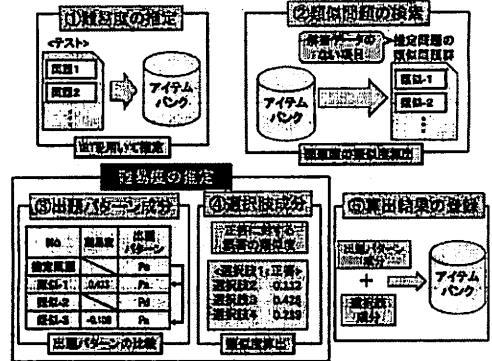


図2 難易度算出手順

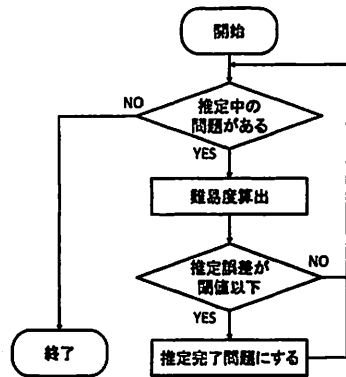


図3 類似問題群ごとの難易度算出のフローチャート

問題である。したがって、選択肢は専門用語となるため、問題1の形式はFである。一方、問題2ではトランスポート層の説明として正しい選択肢を選ばせる問題である。そのため、選択肢はトランスポート層に関する文章となるため、問題2の形式はAである。

具体的な類似度の算出の方法は、まず、各問題の問題文又は選択肢の専門用語（以下、検索語）を含むアイテム・バンク内の問題を検索し、さらに、その問題に含まれる他の専門用語（以下、関連語）を抽出する。そして、各選択肢をそれらの関連語集合を特徴量としたベクトルで表し、選択肢間の類似度を算出する。

関連語の抽出は、アイテム・バンク内の検索語が含まれる問題（以下、関連問題）から行う。しかし、選択問題では検索語と直接関連しない選択肢が存在する場合がある。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、本研究では関連問題内の検索

表 2 検索語の出現箇所に基づく関連語の抽出箇所

出現箇所	正答・誤答	抽出箇所
選択肢	正答	問題文+正答
	誤答	出現した選択肢
問題文		問題文+正答

語の出現箇所(問題文・正答・誤答)により、予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アプローチをとる。

5. 選択肢間の類似度算出手法の検討

5.1 関連語の抽出

表 2 に関連問題の出題パターンが (+) の場合の、関連語の抽出箇所を示す。なお、選択肢に検索語しか含まれない場合は、抽出の対象外とする。また、関連問題の出題パターンが (-) の場合は、正答と誤答を入れ替えた箇所から抽出を行う。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」[11]を用いる。図 4 に関連語の抽出手順を示す。まず、類似度を算出する問題内の検索語を決定する(図 4 中①)。検索語とする専門用語は問題の形式で異なる。以下に、各問題の検索語について説明する。

形式 F

各選択肢は専門用語であるため、選択肢間の類似度は、正答選択肢となる専門用語に対して、誤答選択肢の専門用語がどの程度意味的に似ているのかによって算出する。したがって、各選択肢の専門用語を検索語として関連語を抽出し、その関連語の集合を比較することで正答選択肢に対する誤答選択肢の類似度を算出する。

形式 A

各選択肢は、専門用語に関する説明文である。そのため、選択肢間の類似度は選択肢の内容がどの程度対象知識について説明しているのかによって算出する。したがって、対象知識を検索語として関連語を抽出し、その関連語の集合と各選択肢に含まれる専門用語の集合を比較することで、対象知識に対する選択肢の類似度を算出する。

次に、アイテム・バンク内の検索語が含まれる関連問題を抽出し(図 4 中②)、表 2 の関係性を基に検索語の出現箇所に対応する関連語の抽出箇所を決定する(図 4 中③)。最後に、決定された抽出箇所から関連語を抽出する(図 4 中④)。

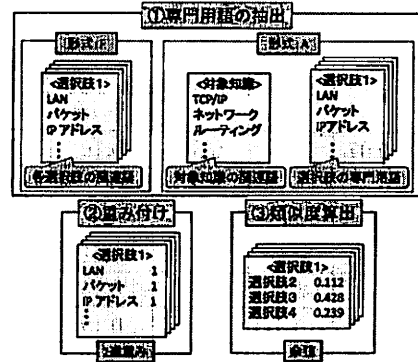


図 4 関連語の抽出手順

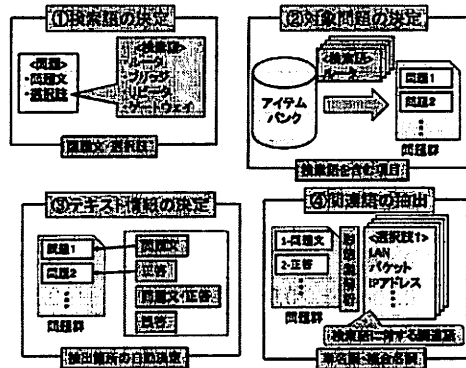


図 5 選択肢間の類似度算出手順

5.2 選択肢間の類似度算出手順

図 5 に、5.1 節で述べた関連語の抽出手順を基にした選択肢間の類似度算出手順を示す。

まず、5.1 節の手順により、各検索語に対する関連語を抽出する(図 5 中①)。形式 F の問題は選択肢ごとに関連語を抽出するため、合計 4 つの関連語の集合が形成される。一方、形式 A の問題では、問題の対象知識に対する関連語と、各選択肢に含まれる専門用語を抽出する。そして、抽出された語に対して重み付けを行う(図 5 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、抽出された語の出現頻度を基に重み付けを行い、出現頻度が 1 の関連語に対しては重み 1 を付与し出現頻度が 2 以上の関連語に対しては重み 2 を付与する。最後に余弦[12]により類似度を算出する

6. 実験・評価

6.1 実験概要

本章では、5 章で検討した選択肢間の類似度の妥当性を検証するために、テストを実施し、解答データと選択肢間の類似度の比較実験を行った。実験には、初級システムアドミ

表 3 形式 F の実験結果

問題	P 値	F 値	正答率	分散	
				類似度	選択率
1	0.07	7.06	0.34	0.10	0.01
2	0.02	16.78	0.42	0.25	0.01
3	0.19	3.05	0.27	0.00	0.00
4	0.23	2.50	0.57	0.11	0.04
5	0.12	4.35	0.51	0.13	0.03
6	0.04	10.31	0.41	0.19	0.02
7	0.02	18.31	0.30	0.10	0.01
8	0.10	5.29	0.39	0.06	0.01
9	0.32	1.79	0.69	0.16	0.09
10	0.04	10.81	0.30	0.06	0.01

表 5 形式ごとの相関係数

形式	相関係数			
	類似度 正答率	選択率 正答率	類似度 選択率	P 値 正答率
F	0.476	0.951	0.352	0.721
A	0.106	0.961	0.242	0.507

ニストレータ試験と基本情報技術者試験で過去に出題された問題 20 問を対象とし、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」を履修している学生 47 名の解答データを用いる。なお、形式 A・F の問題をそれぞれ 10 問ずつ出題する。関連語に抽出では、初級システムアドミニストレータ試験に '04 年度から '08 年度までに出題された問題、および、基本情報技術者試験に '07 年度から '10 年度までに出題された問題 1079 問を対象とし、5.2 節で述べた手順により、選択肢間の類似度を算出した。評価指標としては、まず、類似度の分散と正答率の相関係数を求め、また、類似度と各選択肢が選択される確率（以下、選択率）の相関係数を使用する。

さらに、選択率と類似度の分散比の比較を行った。分散比の比較には F 検定を用い、有意水準 5%、F 境界値は 9.27 として検定を行う。また、P 値と正答率の相関係数を求める。なお、類似度の分散を求める際、選択率の分散と比較しやすくするために、予め選択肢の類似度の合計が 1 になるように正規化を行った。

6.2 実験結果

表 3、表 4 に形式ごとの実験結果、表 5 に形式ごとの類似度の分散、正答率、選択率の分散の相関係数を示す。

類似度の分散と正答率の相関係数は、形式 F で 0.476、形式 A で 0.106 となり、形式 F で

表 4 形式 A の実験結果

問題	P 値	F 値	正答率	分散	
				類似度	選択率
1	0.41	1.33	0.45	0.05	0.03
2	0.10	5.43	0.53	0.01	0.04
3	0.37	1.54	0.47	0.04	0.02
4	0.09	5.85	0.36	0.10	0.02
5	0.36	1.55	0.36	0.02	0.01
6	0.41	1.32	0.74	0.08	0.11
7	0.37	1.53	0.60	0.10	0.06
8	0.18	3.22	0.42	0.00	0.01
9	0.05	9.70	0.37	0.12	0.01
10	0.47	1.11	0.56	0.05	0.05

は類似度の分散と正答率にやや関係性がみられた。また、類似度の分散と選択率の分散の相関係数は形式 F で 0.352、形式 A で 0.242 となり、あまり関係性はみられなかった。

F 検定の結果では、ほとんどの問題において P 値は有意水準 5% を越え、F 値は F 境界値より小さくなり、類似度と選択率の分散が一致することが分かった。また、P 値と正答率の相関係数は形式 F で 0.721、形式 A で 0.507 となり、正答率が低い問題ほど P 値が小さくなった。しかし、形式 F の問題 2, 6, 7, 10 と形式 A の問題 9 ではこの基準を満たさず、分散比が一致しない検定結果となった。そのため、基準を満たさなかったこれらの問題と、P 値が小さい形式 A の問題 2, 4 を除いて再度、類似度の分散と正答率の相関係数を求めた結果、形式 F では 0.857、形式 A では 0.84 となり強い関係性がみられた。

6.3 考察

実験の結果から、類似度と正答率に関係性がある事が分かった。これは、類似度の分散の値が小さいほど、各選択肢は意味的に類似しており、学習者の解答が各選択肢に分散したためだと考えられる。そのため、予めアイテム・バンクに蓄積されている問題群から選択肢間の類似度を算出することで、類似度を基に難易度の推定が可能であると考えられる。

また、P 値と正答率の相関係数から、正答率が低いほど P 値が小さくなる事が分かった。さらに、合計得点が平均点以上の学生を対象とした場合の、類似度の分散と正答率の相関係数は形式 F で 0.604、形式 A で 0.151 となり、全ての学生を対象とした場合より高くなった。これは、難易度の高い問題では、学習者が各選択肢について十分理解できていないため、

類似度がうまく機能しなかったのだと考えられる。そのため、類似度を基に難易度を推定する場合、対象とする問題がどの程度の難易度になるのかを考慮する必要があると考えられる。

しかし、F 検定の結果では複数の問題で類似度の分散と選択率の分散比が一致しない結果となった。これは、今回対象としたテストは本学工学部の3年生が対象の中心となっているため、問題によってはまだ学習していない内容の可能性がある。そのため、複数の問題で分散比が一致しなかったのだと考えられる。

7. まとめと今後の課題

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の推定を目的とし、選択肢の内容を考慮した難易度推定手法を確立するために、選択肢間の類似度について検討を行った。類似度算出の基本的なアイデアは、アイテム・バンク内の問題から各選択肢に関連する用語を抽出し、その関連語集合間の類似度を算出するものである。また、類似度の妥当性の検証実験から、選択肢間の類似度と問題の難易度に関係性がある事が分かった。

今後は、類似度算出の際の関連語の重み付けや不要語の削除方法などについて検討を行い、また、今回F検定におけるP値の低かった問題を分析することで、より正確に選択肢間の類似度の算出を目指す。さらに、算出した類似度をパラメータに含む難易度算出式についての検討を行うことで、より精度の高い難易度の推定を目指す。

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] 近藤彰幸, 濱本和彦, 尾崎将範, 佐藤実, 野須潔: 項目反応理論を応用した信頼性のあるオンライン学習システムの開発, 電子情報通信学会技術研究報告, Vol.106, No.364, pp.13-17 (2006).
- [6] Baker, F. B.: The Basics of Item Response Theory .2nd ed. USA, ERIC Clearinghouse on Assessment and Evaluation, (2001).
- [7] 池田信一, 高木輝彦, 高木正則, 勅使河原可海: 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, 情報処理学会マルチメディア・分散・協調とモバイル(DICOMO2011) シンポジウム, 同シンポジウム論文集, pp.1107-1114 (2011).
- [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] 津森伸一, 海尻賢二: 理解状況に適応した多肢選択式問題の自動生成に関する構想, 教育システム情報学会研究報告, Vol.21, No.4, pp.3-8 (2006).
- [10] 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).
- [11] 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.
- [12] 徳永健伸: 情報検索と言語処理, 東京大学出版社 (1999).