

項目反応理論を用いた類似問題群からの 適応的なテスト出題方式に関する検討

池田 信一[†] 高木 輝彦[‡] 高木 正則^{††} 勅使河原 可海[†]

[†]創価大学工学部 [‡]電気通信大学大学院情報システム学研究科 ^{††}岩手県立大学ソフトウェア情報学部

1 はじめに

近年, ICTを活用した作問学習や作問支援に関する研究が盛んに行われている. また, 様々な教育環境を対象とした作問システムが開発され, 教育現場での実践から作問による学習効果が報告されている[1]. しかし, 作問システムでは, 講義を重ねるごとに問題が増加し, それらの問題の種類や内容の把握・管理が困難になる. また, 多くの問題は再利用されずに埋もれている. そのため, 過去の講義で作成された問題を有効に利用する方法の検討が必要となる.

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方法が提案されている[2]. これは, 問題間の類似度を定量的な数値により算出する手法である. さらに, 高木ら[3]が開発した作問システム「CollabTest」で作成された問題を対象とした評価実験から, 類似度算出の精度の向上が示唆された. 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. この知識とは分野特有の概念や法則, 人物, 歴史など様々であるが, 本研究では作問者が問題内で問うている専門用語とする.

一方で, 学生による類似問題の作成が可能なシステムの開発が行われた[3]. 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, 同一問題を出題するのではなく, 類似問題をランダムに出題することにより, 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が示唆された. しかし, 類似問題をランダムに出題すると, 学習者にとって易しすぎ, また難しすぎる問題が出題される可能性がある. このような問題を解くことによる学習効果は少ないため, 学習者の理解状況に適した難易度の問題の出題が望まれる.

そのため, 本研究では, 問題が類似問題群に分類されていることを前提とし, 作成された問題の有効利用を目的として, 類似問題群からの適応的なテスト出題方式を提案する. 学習者の理解状況に適した難易度の問題を出題するためには, 学習者の理解度と問題の難易度を定量化する必要がある. 本稿では, まず問題の難易度の定量化方法について検討し, その妥当性を検証した結果について述べる.

なお, 本研究では選択問題を対象とする.

2 適応的なテスト出題方式

図1に類似問題群からの適応的なテスト出題の流れを示す. 学習者がテストを要求すると(図1中①), 予め蓄積されている類似度を基に各問題の類似問題を自動で検索し(図1中②), そこから, 学習者に適した難易度の問題を提示する(図1中③).

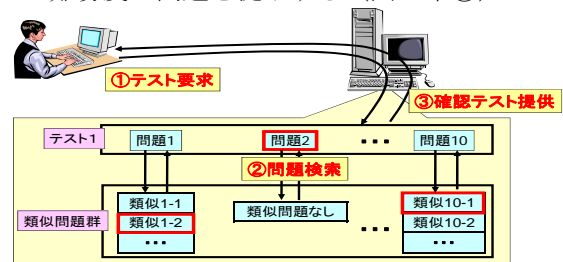


図1 類似問題群からの適応的なテスト出題

これにより, 問題の有効利用が期待される. また, 知識に対して異なる形式や難易度の問題を解くことで多角的な学習が可能となり, より深い理解の定着が期待できる.

本研究では, 学習者の理解度と問題の難易度の定量化に適応型テストで一般的に使用されている項目反応理論 (IRT) [4]を適用する. IRTでは, 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, 本研究では最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を適用する.

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

ここで, D は $D=1.7$ となる定数である. a_j 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. b_j は難易度母数であり, 難易度の値が大きいくほど難しい問題となる. 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 , 問題の識別力母数 a と難易度母数 b を推定することが可能である. 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. しかし, これまでに蓄積されている何万問もの問題を解答させることは困難であるため, テストに出題されていない問題の母数の推定が技術的な課題となる.

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, 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. 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]. そこで, 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する. そして, 分析結果から解答履歴のない問題の難易度の推定を検討する.

A Study on an Adaptive Test Creation Method from the Group of Similar Questions based on IRT

Shinichi Ikeda[†], Teruhiko Takagi[‡], Masanori Takagi^{††}, Yoshimi Teshigawara[†]

[†]Faculty of Engineering, Soka University

[‡]Graduate School of Information Systems, The University of Electro-Communications

^{††}Faculty of Software and Information Science, Iwate Prefectural University

3 難易度推定の検討

3.1 出題パターンと難易度の分析

類似問題において出題パターンの変化に伴い問題の難易度にどのような差が生じるのかを分析する。今回は対象とする問題数の関係上、正しいものを選ばせる問題か、誤ったものを選ばせる問題かの分類は行わず、Pa~Pfの6つの出題パターンで分析を行った。分析方法は、'04年度から本学でCollabTestを利用している講義「コンピュータネットワーク論」において'06年度~'08年度までに作成された問題を対象とし、各年度約80名の解答結果を基に各問題の難易度を式(1)により算出する。その中から、類似問題間で難易度が比較可能な問題49問を抽出し、出題パターンの異なる類似問題間で難易度の差を算出した。表1に2つの出題パターンの組み合わせ全15通りごとの、難易度の差の平均値を示す。

分析の結果、PdやPcが比較的難易度が高かった。これは、PdやPcが専門用語に関連する事柄を問う問題であり、知識を応用して解く必要があるためだと考えられる。

表1 難易度の差の平均

出題パターンの組み合わせ	Pa-Pb	Pa-Pc	Pa-Pd	Pa-Pe	Pa-Pf	Pb-Pc	Pb-Pd
差の平均	0.024	-0.145	-0.471	0.248	-0.018	0.022	-0.535
Pb-Pe	Pb-Pf	Pc-Pd	Pc-Pe	Pc-Pf	Pd-Pe	Pd-Pf	Pe-Pf
	-0.333	-0.384	0.261	0.311	0.721	0.291	-0.539

3.2 出題パターンに基づく難易度推定

前節の分析結果から、難易度を推定する問題（以下、推定問題）とすでに難易度の推定してある類似問題（以下、比較問題）の出題パターンの組み合わせを基に難易度を推定する。具体的には、出題パターンの組み合わせごとに重みを付け、難易度の差の度合いを設定する。次に、推定問題*i*の難易度*PD*を、*n*間の比較問題を用いて推定する流れを説明する。

$$PD = \frac{1}{n} \sum_{j=1}^n (sw_{ij} + d_j) \quad (2)$$

式(2)に本研究で提案する難易度算出式を示す。ここで、*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。次章の実験では、*w_{ij}*は表1を基に-4~4で設定し、*s*は表1の最大値を5段階に分割し、値を0.144とした。また、*d_j*は(1)式により算出した比較問題*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。

4 実験・評価

3.2節で検討した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。3.1節で分析した問題を対象とし、IRTと、3.2節で検討した難易度算出式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2に出題パターンごとの実験結果、表3に比較問題数と誤差を示す。結果から、誤差の

表2 出題パターンごとの結果

出題パターン	比較個数平均	難易度平均本方式	IRT	誤差平均	データ数
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3 比較項目数ごとの誤差

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

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、比較数が多いほどIRTで推定した難易度と近い値となった。しかし、IRTで推定した難易度は一般的に-3~3の値となるため、この誤差は無視できる値ではない。

選択問題では、明らかに誤りだとわ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。今回、出題パターンによる問題の難易度の変化に着目し、選択肢の内容を考慮しなかったため、IRTと本方式で推定した難易度に差が生じたと考えられる。

5 まとめと今後の課題

本研究では、学生が作成した問題の有効利用を目的として、類似問題群から適応的に問題を出題するテスト出題方式を検討している。本稿では、IRTでは推定できない問題の難易度の推定について検討し、比較実験を行った。その結果、出題パターンに基づいた難易度の推定によりIRTで推定した難易度に近い値が推定できることが分かった。今後は難易度の誤差が大きい問題の分析を行い、誤差の原因を明確にし、適した重みの設定方法について検討を行うことで、難易度推定の精度向上を目指す。また、難易度や重みの値を動的に修正する方式を考案し、より汎用性のある推定手法の検討を行う。

参考文献

- [1] 高木正則, 田中充, 勅使河原可海: 学生による問題作成およびその相互評価を可能とする協調学習型 WBT システム, 情報処理学会論文誌, Vol.48, No.3, pp.1532-1545 (2007).
- [2] 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).
- [3] 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.
- [4] 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)
- [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)