

錯乱肢の生成による英語多肢選択問題の適応的難易度調整 Adaptive Difficulty Adjustment of English Multiple-Choice Questions by Generating Distractors

内田 時央¹⁾ バーゴ アンドリュウ¹⁾ 黄瀬 浩一¹⁾
Tokio Uchida Andrew W. Vargo Koichi Kise

1 はじめに

近年, ICT を活用した新たな学習形態として Adaptive Learning が注目を集めている. Adaptive Learning とは学習者の能力や特性に適した問題を自動で提示するシステムであり, 事前に用意した問題から推薦する方式が主流となっている [1, 2]. しかし, この方式では既存の問題のみ用いて学習を行うため, 学習者個人に対して完全に適合することは難しいと考えられる. そこで本研究では, 学習者に合わせて新たな問題を生成することで真に個人適合可能な Adaptive Learning の構築を考える.

本研究は空所補充型の英語多肢選択問題を扱う. この形式の問題は英検や TOEIC など様々な試験で用いられており, 記述形式の問題と比べて解答に要する時間が短いため効率よく学習を進めることができる. また任意の英文に対して空所と選択肢を与えることで新たな問題が生成できるため, 学習者が利用したい英文を用いた問題を提示することが可能となり, 高い学習効果を生むと考えられる. 本研究では, 学習者に合わせて適切な英文を選択し, その英文から新たな多肢選択問題を生成することで真に個人適合可能なシステムの構築を目指す.

英文から多肢選択問題を生成する際, その問題の難易度は誤答選択肢である錯乱肢が大きく影響する [3]. そのため能力の異なる学習者が同じ問題文を用いて学習を行う場合, それぞれの学習者に合わせて錯乱肢を生成する必要がある. 本研究では, 真に個人適合可能なシステムを構築するための最初の段階として, 既存の問題から新たな錯乱肢を生成することで学習者に合わせた問題の難易度調整を行う手法を提案する.

錯乱肢の生成により問題の難易度調整を試みた研究として, 湯浅らの研究 [4] がある. 湯浅らは Masked Language Model (MLM) と呼ばれる文章中の隠された単語を推定するモデルを用いて, 各単語が推定される順位に基づき錯乱肢を変化させることで難易度を調整している. しかし, この研究の課題として, 正解となる単語や明らかに誤っている単語が錯乱肢に含まれることで不適切な錯乱肢が学習者に提示される場合がある.

本研究では, 不適切な錯乱肢を取り除いて難易度調整を行うために Replaced Token Detection (RTD) を利用する. RTD は文章中の各単語に対して誤って置換された単語か否か識別するモデルである. このモデルが各単語に対して出力する数値を用いることで, 適切に難易度調整を行うことができると考えられる. 日本人 11 名 (性別: 男性 11 名, 平均年齢 21.8±1.6 歳, TOEIC: 732.5±114.7 点) を対象にした実験の結果, 一部の参加者は提案手法によって適切な難易度を維持した学習が可能になると確認された. また大半の参加者は元の問題の

1) 大阪公立大学大学院 情報学研究科

Graduate School of Informatics, Osaka Metropolitan University

錯乱肢を用いた学習時の結果と比べて平均正解率が好転していたことから, 提案手法による難易度調整が適切に行われたことが示唆された. なお, 本研究は大阪府立大学大学院工学研究科倫理委員会の承認を得ていることを付記しておく.

2 関連研究

2.1 既存の Adaptive Learning

Ruan らの研究 [1] では科学知識や英単語などを学習するための対話型エージェントとして QuizBot を提案している. QuizBot では事前にクラウドワークによって難易度が付与された問題を用意し, 提示した問題に対する学習者のフィードバックを対話形式で取得することで次に提示すべき問題を選定している. また Sfenrianto らの研究 [2] では学習者の能力に合わせた英語教材を推薦するシステムとして Adaptive Learning Systems-Knowledge Level (ALS-KL) を提案している. ALS-KL では学習者の能力と問題の難易度をそれぞれ初級・中級・上級の 3 段階に分類し, プレテストの結果をもとに学習者が利用すべき問題を選定している.

このように現在の Adaptive Learning では事前に用意した問題から推薦する方式が主流となっている. しかし, この方式では既存の問題のみ用いて学習を行うため, 学習者個人に対して完全に適合することは難しいと考えられる. 本研究では, 学習者に合わせて新たな問題を生成することで真に個人適合可能な Adaptive Learning の構築を目指す.

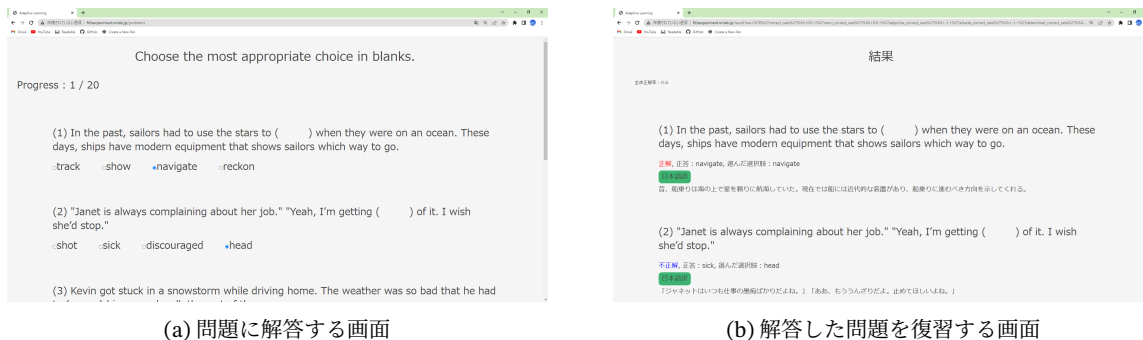
2.2 錯乱肢が難易度に及ぼす影響

Susanti らの研究 [3] では語彙を問う英語多肢選択問題に関して難易度に影響する要素を調査している. 語彙を問う英語多肢選択問題とは, 本文中の特定の単語に最も意味が近い選択肢を尋ねる形式の問題である. この研究では問題の構成要素を問題文の読解難易度, 正解と錯乱肢の類似度, 錯乱肢の単語自体の難易度の 3 つに分類し, 各要素が問題の正解率に与える影響を検証している. その結果, 正解と錯乱肢の類似度および錯乱肢の単語自体の難易度の 2 つが正解率に大きな影響を与えることが示されている.

錯乱肢が難易度に及ぼす影響は空所補充型の英語多肢選択問題の場合も同様に大きいと考えられる. 本研究では錯乱肢の生成により難易度調整を行うことを目指す.

2.3 錯乱肢の自動生成

空所補充型の英語多肢選択問題に関して錯乱肢の自動生成を行った研究として Panda らの研究 [5] や Chen らの研究 [6] がある. Panda らは問題文の空所に正解を当てはめた文章に対して機械翻訳を繰り返して行い, その過程で得られた正解と意味の近い単語を錯乱肢として生成している. また Chen らは正解単語の品詞の種類に基づいて活用形や語頭・語尾を変化させることで新たな錯乱肢を生成している.



(a) 問題に解答する画面

(b) 解答した問題を復習する画面

図 4: 実験で用いた Web アプリケーションの画面

RoBERTa [7] といった言語モデルの事前学習方法を改良することで、同等サイズの他モデルよりも高い性能を出している。本研究では Hugging Face¹⁾で公開されている事前学習済みの ELECTRA を利用した。

4.2 モデルの追加学習

本研究では事前学習済みの RTD に対して、英語の専門家により質が評価された空所補充型の英語多肢選択問題に関するデータセット²⁾から得られた 3167 問を用いて、錯乱肢として明らかに誤っている単語を追加学習させた。このことにより RTD がより高い精度で不適切な錯乱肢を識別できるようになることを期待した。

4.3 不適切な錯乱肢を取り除くための閾値

本実験では湯浅らの研究 [4] で見られた不適切な錯乱肢に着目し、閾値 δ と θ をそれぞれ 0 と 10 と定めた。

4.4 難易度調整の基準

本実験ではシステムによる難易度調整の基準として学習中の平均正解率を用いた。その詳細として、まず適切な正解率を調査した研究 [10] の結果をもとに学習中における適切な難易度を平均正解率が 60% から 90% の間にある状態と定義した。そのうえで参加者が問題を 5 問解くごとに問題全体の平均正解率を計算し、その正解率が 60% 以下の場合には錯乱肢を選択する範囲を 10 単語分だけ下に移動させて問題を易化させ、90% 以上の場合には範囲を 10 単語分だけ上に移動させて問題を難化させた。

4.5 実験に用いた問題

学習中の平均正解率をもとに問題の難易度調整を行ううえで学習に用いる問題の元々の難易度が統一されている必要がある。そこで本実験では、英検の過去問を用意して 2 通りの学習に用いる全問題の級を統一することで、各問題間で生じる難易度差を軽減した。本実験は英検 1 級、準 1 級、2 級の過去問を用意して実施した。

4.6 実験環境および実験条件

本実験は、Flask により開発した Web アプリケーションを用いて、各参加者が所有するパソコン上で行った。参加者は図 4(a) に示すような多肢選択問題を解答し、5 問解答し終わるごとに図 4(b) に示すような解答の正誤、正解の選択肢、学習者が選んだ選択肢、問題文の日本語訳を得た。参加者にはタスクとして 1 日につき 1 時間で最低 100 問の学習を数日間行ってもらうように指示し、その謝礼に 1 時間の学習につき 1,000 円分の金券を支払った。

5 結果・考察

5.1 各学習方法における平均正解率の変化

全参加者が解答した英検準一級の過去問を使用した場合の、提案手法による学習時と元の問題の錯乱肢による学習時の平均正解率の変化を確認する。図 5(a) および図 5(b) はそれぞれ、英検準 1 級の過去問を使用した場合の、提案手法による学習時と元の問題の錯乱肢による学習時の平均正解率の変化を示している。また図中の背景が赤い部分は本実験で適切な難易度と定義した平均正解率の範囲を表している。ただし図 5(b) に関して、ネットワークトラブルで正しくデータを取得できなかった p02 は含まれていない。

図 5(a) より、参加者 11 名中 7 名 (p01, p03, p05, p06, p08, p10, p11) は提案手法で学習した時の最終的な平均正解率が 60% から 90% の間にあることから、多くの参加者は提案手法によって適切な難易度を維持した学習が可能になることが確認された。また図 5(b) より、参加者 11 名中 2 名 (p10, p11) のみが元の問題の錯乱肢で学習した時の最終的な平均正解率が 60% から 90% の間にあり、p01, p03, p05, p06, p08 の 5 名に関しては提案手法で学習した時のみ適切な難易度を維持した学習へと推移したことから、提案手法による問題の難易度調整が適切に行われたことが示唆された。

5.2 錯乱肢の単語自体の難易度に関する考察

英検準 1 級を用いた場合の提案手法による学習に関して、学習中の問題の難易度変化に対する印象を尋ねた設問の回答結果を図 6 に示す。図 6 より、実験結果では参加者 11 名中 7 名が適切な難易度を維持した学習へと推移したにも関わらず、11 名中 7 名が「難易度が変化したとは感じず、常に難しかった。」と回答していた。そのうち参加者 5 名 (p01, p03, p06, p08, p10) の回答理由を表 1 に示す。表 1 より、常に問題が難しいと感じた理由として、生成された錯乱肢の単語自体の難易度が高かったことが一因と考えられる。そのため今後の課題として、錯乱肢の難易度を定める基準に RTD の出力だけでなく単語の frequency 等も考慮する必要がある。

6 結論

本論文では、真に個人適合可能な Adaptive Learning を構築するための最初の段階として、既存の多肢選択問題から新たな錯乱肢を生成することで学習者に合わせた問題の難易度調整を行う手法を提案した。さらに、提案手法による難易度調整が学習者に合わせて適切に行われるかを検証するための実験を行った。その結果、提案手法によって多くの参加者は提案手法によって適切な難易

1) <https://huggingface.co/>

2) 早稲田大学 理工学術院 英語教育センター (CELESE) の Ralph Rose 教授から提供していただいたデータセット

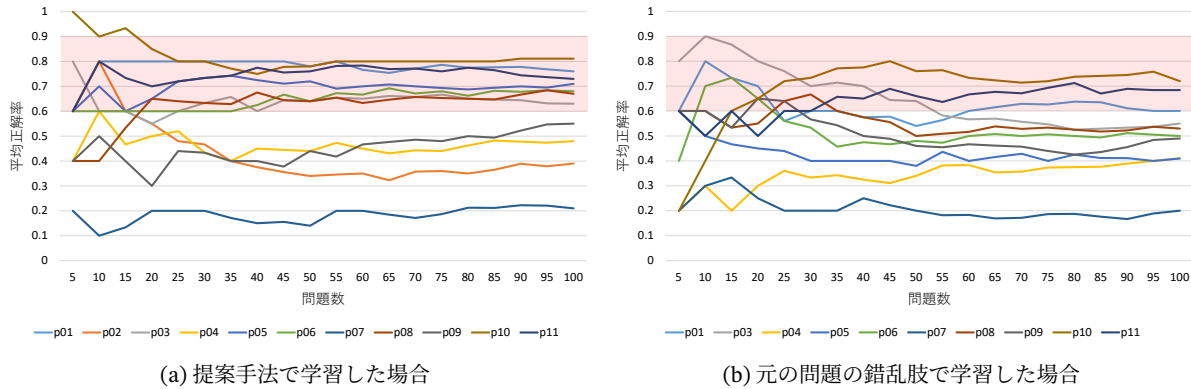


図5: 英検準1級を用いた場合の平均正解率の変化

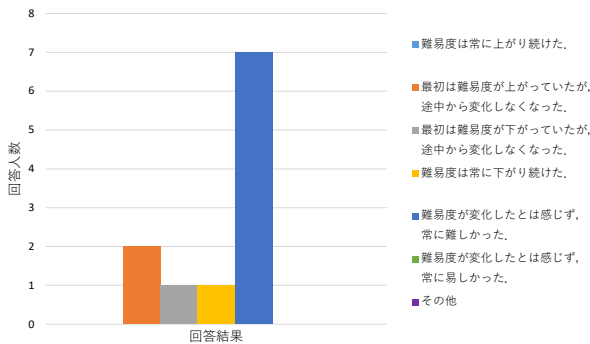


図6: 英検準1級の過去問を提案手法で学習した際の難易度変化に対する印象

表1: 難易度が常に難しいと感じた理由

参加者番号	回答理由
p01	ずっと難しい問題があったように思えたから。
p03	体感で、途中から問題を解くスピードが落ちた様に思うため
p06	そもそも単語の語彙がわからないものが多かった
p08	単語が難しかった
p10	常に、自分の知らない単語の割合が同じであった。

度を維持した学習が可能になることが確認された。また大半の参加者は提案手法で学習した時のみ適切な難易度を維持した学習へと推移したことから、提案手法による問題の難易度調整が適切に行われたことが示唆された。今後の課題として、提案手法における不適切な錯乱肢を取り除く方法について見直すことや難易度調整の基準についても見直すこと、そして錯乱肢以外の要素も踏まえた難易度調整の手法を考えること等が挙げられる。

謝辞

本研究の一部は JST Trilateral AI Research (JP-MJCR20G3), JSPS 科研費基盤 (B)(20H04213), JSPS 国際共同研究強化 (B)(20KK0235), 阪大 Society5.0 グランドチャレンジの補助による。

参考文献

[1] Sherry Ruan, Liwei Jiang, Justin Xu, Bryce Joe-Kun Tham, Zhengneng Qiu, Yeshuang Zhu, Elizabeth L. Mur-

nane, Emma Brunskill, and James A. Landay. Quizbot: A Dialogue-Based Adaptive Learning System for Factual Knowledge. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–13, Association for Computing Machinery, 2019.

[2] Sfenrianto Sfenrianto, Yustinus B. Hartarto, Habibullah Akbar, Mukhneri Mukhtar, Efriadi Efriadi, and Mochamad Wahyudi. An Adaptive Learning System based on Knowledge Level for English Learning. International Journal of Emerging Technologies in Learning (IJET), Vol. 13, No. 12, pp. 191–200, 2018.

[3] Yuni Susanti, Takenobu Tokunaga, Hitoshi Nishikawa, and Hiroyuki Obari. Controlling item difficulty for automatic vocabulary question generation. Research and Practice in Technology Enhanced Learning, Vol. 12, No. 1, p. 25, 2017.

[4] 湯浅 成章, Andrew Vargo, 黄瀬浩一. 個人に適した英語多肢選択問題の自動生成方法の提案. 情報処理学会研究報告教育学習支援情報システム (CLE), 第 2022-CLE-36 巻, pp. 1–8. 2022.

[5] Subhadarshi Panda, Frank Palma Gomez, Michael Flor, and Alla Rozovskaya. Automatic Generation of Distractors for Fill-in-the-Blank Exercises with Round-Trip Neural Machine Translation, In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pp. 391–401, Association for Computational Linguistics, 2022.

[6] Chia-Yin Chen, Hsien-Chin Liou, and Jason S. Chang. FAST - An Automatic Generation System for Grammar Tests, In Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions, pp. 1–4, Association for Computational Linguistics, 2006.

[7] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A Robustly Optimized BERT Pretraining Approach, 2019

[8] Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. ELECTRA: Pre-Training Text Encoders as Discriminators Rather Than Generators, In Proceedings of ICLR 2020, 2020

[9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Association for Computational Linguistics, pp. 4171–4186, 2019.

[10] Robert C Wilson, Amitai Shenhav, Mark Straccia, and Jonathan D Cohen. The Eighty Five Percent Rule for optimal learning. Nature Communications, Vol. 10, No. 1, pp. 1–9, 2019.