



2022年8月31日

報道機関 各位

東北大学大学院情報科学研究科

### 人間と AI の協調により記述式答案自動採点の品質を保証 — AI による安全な自動採点の実現に道 —

#### 【発表のポイント】

- 深層学習（注1）の登場により人工知能（AI）による記述式答案の自動採点精度は飛躍的に向上しており、実用化への機運が高まっている。
- AI による採点誤りの懸念は依然として存在しており、AI による自動採点品質を保証することが今後の実用化に向けて大きな課題となっている。
- 人間の採点者と AI が自動採点結果の信頼度をもとに採点作業を分担することで採点の品質を保証する実践的な枠組みを構築し、その実現性を確かめた。
- 本研究を足がかりにして、AI による自動採点の品質保証に向けた研究開発がさらに進むことが期待される。

#### 【概要】

深層学習を用いた機械学習手法の登場により、AI による記述式答案の自動採点精度は著しく向上しています。一方で、学習データに現れたことのない新しい表現を含む答案に対して適切に採点を行うことは依然として難しいため、AI が予期しない採点誤りを犯すことへの懸念が自動採点の実用化に向けての大きな課題の一つとなっています。

東北大学大学院情報科学研究科/理化学研究所の舟山弘晃研究員（博士後期課程）、佐藤汰亮研究員（現在は楽天グループ株式会社所属）、松林優一郎准教授、鈴木潤教授、乾健太郎教授、フューチャー株式会社 /理化学研究所の水本智也研究員（現在はLINE 株式会社所属）らの研究グループは、この課題の解決に向けて、人間の採点者と AI が採点タスクを分担することで、採点品質を保証するための実践的なフレームワークの構築を試み（図1）、

このフレームワークにより採点品質を適切にコントロールできることを明らかにしました。

本研究成果は、英国ダラム大学で開催された教育に関する人工知能の国際会議 The 23rd International Conference on Artificial Intelligence in Education (AIED2022) で採択され、7月27日に発表いたしました。

### 【詳細な説明】

深層学習を用いたアプローチやオンライン百科事典 Wikipedia のような大規模データで言語知識の学習を行ったモデルの登場により、AI による記述式答案の自動採点精度は飛躍的に高まっています。特に、数十字程度で記述される答案を対象とした短答式記述問題の自動採点では、一部の問題において人間の採点者と同等レベルの採点品質を実現できることが報告されています。しかしながら、採点 AI は学習データ中に存在しないような未知の表現を含む答案に対して、適切に採点を行うことが難しいことが知られており、このような AI による採点誤りへの懸念は、自動採点の実用化の大きな妨げとなっています。

この課題の解決にむけて、自動採点システムと人間が協調的に採点を行う採点フレームワークを構築しました（図1）。

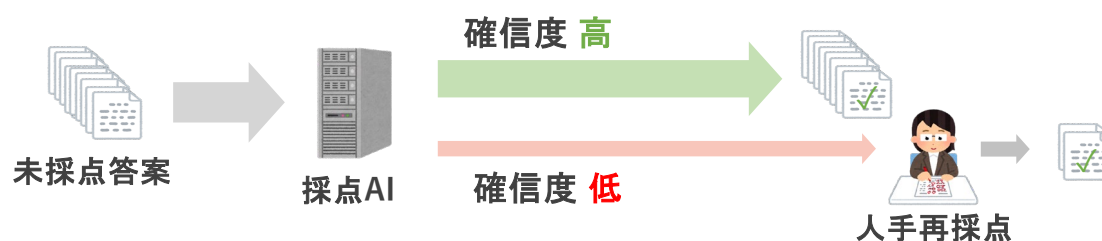


図 1. 人間の採点者と AI の分担による採点品質の保証

この枠組みでは、採点 AI による採点結果の信頼性を表す尺度である確信度を活用します。各答案に対して自動採点結果の確信度を確認し、確信度が低い場合は人間の採点者による再採点を行います。具体的な手続きとしては、まず少量の採点済み答案データをもとに、所望の採点品質を実現するための確信度の下限を推定し、実際の自動採点の際に確信度がその下限値を下回った場合に人間が再採点を行うことで、所望の採点品質の実現を図ります。

本研究では、国内と英語圏の記述式問題のデータセットを用いて、構築したフレームワークが実際の採点データ上で期待通りの効果を示すか確かめるためシミュレーションを行い、その実現性を明らかにしました（図2）。また、人間の採点者間の採点結果の一致率が高い問題ほど、本フレームワーク

を適用することで、より高品質な採点を低コストで実現可能であることを明らかにしました。

本研究は、自動採点を実際の採点現場において導入する上で重要な観点である採点品質の保証を図るための重要な一歩として位置付けられます。本研究で明らかになった知見をもとに、今後自動採点の実用化がさらに進んでいくものと期待されます。

図 2. 目標とする答案全体の採点誤差 (x 軸) と提案するフレームワークで得られる答案全体の採点誤差 (y 軸) の関係。左図は英語圏の採点データ、右図は国内の採点データを用いたときの結果を表す。x 軸、y 軸ともに、Root Mean Squared Error (RMSE)を示しており、値が小さいほど採点品質が高い(採点誤差が小さい)ことを示す。どちらのデータセットにおいても、安定的に目標の採点品質を実現している。



### 【用語解説】

(注 1) 深層学習

人間の脳神経回路を計算機的に模倣したニューラルネットワークを多層に組み合わせる機械学習手法の一つ。近年では、画像認識、言語処理、音声認識など様々な分野で盛んに活用されている。

### 【論文情報】

タイトル: Balancing Cost and Quality: An Exploration of Human-in-the-Loop Frameworks for Automated Short Answer Scoring

著者: Hiroaki Funayama, Tasuku Sato, Yuichiroh Matsubayashi, Tomoya Mizumoto, Jun Suzuki and Kentaro Inui

掲載誌: The 23rd International Conference on Artificial Intelligence in Education (AIED2022)

DOI: 10.1007/978-3-031-11644-5\_38

URL: [https://link.springer.com/chapter/10.1007/978-3-031-11644-5\\_38](https://link.springer.com/chapter/10.1007/978-3-031-11644-5_38)

**【問い合わせ先】**

＜ 研究に関する事 ＞

東北大学 大学院情報科学研究科 担当 乾健太郎  
TEL: 022-795-7091 E-mail: [inui@tohoku.ac.jp](mailto:inui@tohoku.ac.jp)

＜ 報道に関する事 ＞

東北大学 大学院情報科学研究科 広報室 担当 佐藤みどり  
TEL: 022-795-4529 E-mail: [koho@is.tohoku.ac.jp](mailto:koho@is.tohoku.ac.jp)