

令和 4 年 6 月 8 日現在

機関番号：12612

研究種目：挑戦的研究（萌芽）

研究期間：2019～2021

課題番号：19K21751

研究課題名（和文）小論文自動採点実用化のためのルーブリックを考慮した項目反応理論の開発

研究課題名（英文）Automated essay scoring using item response theory that considers rubric characteristics

研究代表者

植野 真臣（Ueno, Maomi）

電気通信大学・大学院情報理工学研究科・教授

研究者番号：50262316

交付決定額（研究期間全体）：（直接経費） 4,900,000円

研究成果の概要（和文）：近年、大学入試や英語4技能試験などの大規模テストにおいて小論文試験のニーズが急速に高まってきているが、その採点コストの大きさが課題となっている。この問題を解決するアプローチの一つとして、自動採点技術の開発が急務とされているが、現時点では実用化レベルの高精度の達成は難しい。一方で、我々は小論文試験において評価者の特性を考慮した高精度なスコアリングを実現する項目反応理論を開発してきた。本研究では、このモデルを拡張して評価者とルーブリックも項目とみなす新しい項目反応理論を開発するとともに、それをを用いて異なる特性を持つ様々な最先端の自動採点機を統合する手法を開発する。

研究成果の学術的意義や社会的意義

既存の自動採点機を一つ一つ深く研究しても現実の小論文自動採点を実現することは難しい。それに対して提案手法は一つの自動採点機を単独で用いるよりも必ず精度を向上させるという利点があり、急務である小論文の自動採点の実用化を大幅に加速すると期待できる。さらに、このアプローチは個々の自動採点機を統合して人間を近似する手法の研究のみに集中できることや、新しく精度の良い自動採点機が提案されても簡単に組み入れることができる、など長期的な利点が多い。自動採点の実用化されれば大学入試への応用だけでなく、高度な支援機能を持つ学習支援システムなどが実現でき、教育界・産業界にもインパクトが期待できる。

研究成果の概要（英文）：In automated essay scoring (AES), essays are automatically graded without human raters. Many AES models based on various manually designed features or various architectures of deep neural networks have been proposed over the past few decades. Each AES model has unique advantages and characteristics. Therefore, rather than using a single AES model, appropriate integration of predictions from various AES models is expected to achieve higher scoring accuracy. In the present paper, we develop 1) a new item response theory model that can estimate scores while considering characteristics of individual human-raters and rubric-items, and 2) a method that uses the item response theory model to integrate prediction scores from various AES models while taking into account differences in the characteristics of scoring behavior.

研究分野：教育工学

キーワード：自動採点 ルーブリック 項目反応理論

1. 研究開始当初の背景

近年、大学入試における小論文試験や英語4技能試験におけるライティング試験など、大規模な小論文テストのニーズが急速に高まっている。このような大規模試験の問題として、評価者による採点のコストが膨大となることが挙げられる。

この問題を解決する方法の一つとして、自動採点技術の導入が検討されている。e-Rater に代表される従来の自動採点技術では、答案文と模範解答文の統計的類似性を用いることにより採点される。しかし、論理的な意味を評価できないため、全く意味が正答とはかけ離れている文章でもキーワードが使われているだけで高い点数になってしまい、日本の入試では用いることができない。これに対し、近年、自然言語処理分野において LSTM などのディープラーニングの最先端技術を用いて文章の論理的な意味を含んで評価できる自動採点技術が提案されてきている。しかし、人間の評価と比較すると誤差が多く、大学入試レベルの小論文を採点できるレベルには至っていない。

一方、応募者らは、小論文試験のようなパフォーマンステストにおいて、人間評価者の特性（甘さ/厳しさの差異や一貫性の欠如など）を考慮した信頼性の高いスコアが推定できる項目反応理論を開発してきた。項目反応理論はテスト理論の数理モデルで、異なるテストを受検しても同一尺度上に精度の高いスコアを得ることができ、すでに情報処理技術者試験や一部の公務員試験、英検や SPI など多くのハイステークテストで採用されている。応募者らの開発した手法は、課題だけでなく評価者も問題項目とみなし、評価者パラメータを持つ項目反応理論を提案している。この手法は評価者のバイアスを取り除いてスコアを推定できるため、素点平均などの単純な手法と比べて高精度なスコアリングが可能となる。すでに医療系大学間共用試験や英検、リクルートキャリア社の人事考課試験など、様々なパフォーマンステストで採用されている。

本研究の主なアイデアはこのモデルを拡張して、人間の代わりに様々な自動採点機をそれぞれ異なる評価者とみなして統合し人間に近似させる自動採点機を開発することである。しかし、各自動採点機は異なるルーブリック（評価基準）をもって評価するのでルーブリック項目への欠測値を推測しながらスコアを算出できる新しいモデルが必要となる。

2. 研究の目的

上記の背景を踏まえ、本研究の目的は次の通りとする。

- (1) ルーブリックの項目パラメータを組み込んだ新しい項目反応理論のモデルを提案する。
- (2) 提案モデルを用いて様々な最先端自動採点機を評価者とみなして人間評価者を近似するように特性パラメータを推定し、未知の受検者の小論文を自動採点機で採点する手法を提案する。
- (3) 実データに当てはめ、提案手法の評価を行う。

本提案の利点は以下が考えられる。(1) 一つだけの自動採点機を用いるより評価の精度が必ず向上する。(2) 人間評価者のスコアに複数の自動採点機のスコアが近似できるようにパラメータ推定できる。(3) 人間の評価者による採点負担を大幅に削減することができる。(4) 個々の自動採点機を分析でき、新しい自動採点機が提案されても容易に追加できる。

3. 研究の方法

目的 (1) において、本研究では課題パラメータ、評価者パラメータのみでなくルーブリック項目パラメータを付け加えた新しい項目反応理論を提案する。このモデルを用いて複数の自動採点機と人間の評価者による評点データからそれぞれ特性を考慮したスコアリングを行う。しかし、単純に自動採点機を評価者の一部とみなしてモデルに適用すると、人間の評価者に対して

自動採点機の数が増加したとき、評価の尺度が人間の評価者と変わってしまう危険性がある。

この問題を解決するために、目的 (2) では、次の手順でモデルの学習を行う。まず、大量の小論文データに対して人間の評価者による採点データのみを用いて各評価者とルーブリック項目の特性パラメータ、個々の小論文のスコア推定を行い、人間評価者のスコア尺度を作成する。そして、尺度を固定したまま、自動採点機による小論文への評価データを加え、各自動採点機の特性パラメータ (評価者・ルーブリック・パラメータ) を推定する。この際、欠測しているルーブリック項目の特性パラメータはベイズ推定 (MCMC) により補完する。推定されたパラメータ値を用いて、未知の受検者の小論文を自動採点機で採点する。以降は、評価対象物に割り当てる人間の評価者を減らして、自動採点機と評価者の採点データを用いてスコアリングを行う。

目的 (3) では、シミュレーション実験と被験者実験を行い、提案モデルの有効性を評価する。被験者実験では、最先端の小論文の自動採点機が英語にしか対応していないため英語の小論文を対象とする。TOEFL や TOEIC などで使用された課題を被験者に出題し、その回答を実試験と同様のルーブリックを用いて複数の人間の評価者に採点させ、同時にそれぞれ異なるルーブリックを持つ自動採点機にも採点させることでデータを収集する。これらのデータに提案モデルを適用し、(A) 既存の自動採点機より高い精度を達成できるか、(B) 人間の評価者を減らしても高精度な採点が維持できるか、(C) 十分に学習が進んだ後には、人間の評価者なしの自動採点機のみで高精度な採点の実現できるか、の評価を行う。

4. 研究成果

(1) ルーブリックパラメータを持つ項目反応理論の開発

評価者パラメータのみでなくルーブリック項目パラメータを付け加えた新しい項目反応理論を開発した。具体的には、ルーブリック評価で得られる 4 相データ (被評価者 × 課題 × 評価者 × 評価観点で構成される) に適用でき、課題・評価者・ルーブリックの評価観点の特性を同時に考慮した被評価者のスコアリングが可能な項目反応モデルを開発した。さらに、ルーブリックを用いた評価では、推定したいスコアに多次元性が仮定される場合がある。そこで、ルーブリックの複数評価観点の背後に構成される潜在的な多次元スコアを推定できる上記のモデルの拡張モデルも開発した。

シミュレーションと実データ実験の結果、これらのモデルによって課題・評価者・ルーブリックの評価観点の特性を適切に推定できるとともに、それらの影響を考慮した高精度なスコアリングが実現できることを確認した。

本研究の成果は、教育工学分野のトップ国際会議の一つである *Artificial Intelligence in Education (AIED, CORE ランク A)* にフルペーパーとして採択され、ベストペーパーにノミネートされた[1]。さらに、査読付き論文は、電子情報通信学会論文誌 *D* に 2 報[2, 3]、国際論文誌 *Behaviormetrika, Springer* に 1 報[4]が掲載された。

(2) 項目反応モデルを用いた自動採点手法

自動採点モデルは性質の多様化が進み、それぞれのモデルは異なる利点を有している。本研究では、多様な自動採点モデルが予測したスコアを平均化することで、スコアの予測精度の向上を目指した。ただし、自動採点モデルの特性が多様であるがゆえに、単純にスコアを平均化するだけでは精度の向上が妨げられる恐れがある。この問題に対する解決策として本研究では、項目反応理論を利用している。4. (1) で述べたように、近年、評価の一貫性や厳しさといった人間評価者の特性を考慮してスコアを推定できるモデルが多数提案されており、高精度なスコア推定が実現されている。本研究では、我々が開発してきた項目反応モデルである一般化多相ラッシュモデルを、自動採点モデルを人間評価者とみなすことで適用し、自動採点の精度向上を図る。提案手法は、各自動採点モデルの特性を考慮しつつ各モデルの予測スコアを統合することができるため、単一の自動採点モデルや単純なスコアの平均化手法と比べてより正確な予測スコアを得ることができる。

また、本研究では、トップジャーナルである IEEE や人工知能・統計的自然言語処理のトップカンファレンスである Association for the Advancement of Artificial Intelligence (AAAI), Association for Computational Linguistics (ACL)などに採録された最先端の自動採点モデルを実装するだけでなく、独自に高精度な自動採点モデルの開発も行なった。具体的には、近年広く研究がなされている深層学習ベースの自動採点アプローチと、古く研究されてきた人手設計した特徴量を用いた自動採点アプローチを統合する技術を開発した。そして、独自開発した自動採点モデルを含む最先端の複数自動採点モデルを用いて、提案手法の性能を実データ実験で評価した。実験の結果、提案手法により、単一の自動採点モデルや単純なスコアの平均化手法と比べて、自動採点の性能が向上することが確認できた。

本研究の成果は、教育工学分野のトップ国際会議の一つである Artificial Intelligence in Education (AIED, CORE ランク A) と自然言語処理分野の主要国際会議の一つである International Conference on Computational Linguistics (COLING, CORE ランク A) に採択された[5, 6]。さらに、査読付き論文が、電子情報通信学会論文誌 D に掲載された[7]。また、人工知能学会の全国大会で成果発表を行い、全国大会優秀賞を受賞した[8]。最終成果は、国際論文誌 IEEE Transactions on Learning Technologies に現在投稿中である。

<引用文献>

- [1] Masaki Uto (2021) A Multidimensional Item Response Theory Model for Rubric-based Writing Assessment. International Conference on Artificial Intelligence in Education (AIED), Lecture Notes in Computer Science, vol.12748, pp.420-432. <Best paper award nominee>
- [2] 宇都雅輝・植野真臣 (2020) ルーブリック評価における項目反応理論. 電子情報通信学会論文誌 D. Vol. J103, No. 05. pp. 459-470.
- [3] 宇都雅輝 (2022) ルーブリックを用いたパフォーマンス評価のための多次元 4 相型項目反応モデル. 電子情報通信学会論文誌 D, Vol. J105-D, No. 07.
- [4] Masaki Uto (2021) A multidimensional generalized many-facet Rasch model for rubric-based performance assessment. Behaviormetrika, Springer, Vol.48, Issue 2, pp.425-457.
- [5] Itsuki Aomi, Emiko Tsutsumi, Masaki Uto, Maomi Ueno (2021) Integration of Automated Essay Scoring Models using Item Response Theory. International Conference on Artificial Intelligence in Education (AIED), Lecture Notes in Computer Science, vol.12749, pp.54-59.
- [6] Masaki Uto, Yikuan Xie, Maomi Ueno (2020) Neural Automated Essay Scoring Incorporating Handcrafted Features. Proceedings of the 28th International Conference on Computational Linguistics (COLING), pp.6077-6088.
- [7] 青見樹・堤瑛美子・宇都雅輝・植野真臣 (2021) 項目反応理論による小論文自動採点機のモデル平均. 電子情報通信学会論文誌 D. Vol. J104-D, No. 11, pp.784-795.
- [8] 青見樹・堤瑛美子・宇都雅輝・植野真臣 (2021) 項目反応理論を用いた自動採点モデルの統合手法. 第 35 回人工知能学会全国大会. (大会優秀賞受賞)

5. 主な発表論文等

〔雑誌論文〕 計12件（うち査読付論文 11件 / うち国際共著 0件 / うちオープンアクセス 5件）

1. 著者名 宇都 雅輝	4. 巻 J105-D
2. 論文標題 ループリックを用いたパフォーマンス評価のための多次元4相型項目反応モデル	5. 発行年 2022年
3. 雑誌名 電子情報通信学会論文誌D 情報・システム	6. 最初と最後の頁 in press
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2021JDP7042	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 青見 樹、堤 瑛美子、宇都 雅輝、植野 真臣	4. 巻 J104-D
2. 論文標題 項目反応理論による小論文自動採点機のモデル平均	5. 発行年 2021年
3. 雑誌名 電子情報通信学会論文誌D 情報・システム	6. 最初と最後の頁 784 - 795
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2021JDP7002	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Uto Masaki、Ueno Maomi	4. 巻 Feb-47
2. 論文標題 A generalized many-facet Rasch model and its Bayesian estimation using Hamiltonian Monte Carlo	5. 発行年 2020年
3. 雑誌名 Behaviormetrika	6. 最初と最後の頁 469 - 496
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s41237-020-00115-7	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 青見 樹、菅原 聖太、植野 真臣	4. 巻 J103-D
2. 論文標題 アンサンブル学習によるモデル平均ベイジアンネットワーク分類器	5. 発行年 2020年
3. 雑誌名 電子情報通信学会論文誌D 情報・システム	6. 最初と最後の頁 183 - 193
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2019JDP7059	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 宇都 雅輝、植野 真臣	4. 巻 J103-D
2. 論文標題 ルーブリック評価における項目反応理論	5. 発行年 2020年
3. 雑誌名 電子情報通信学会論文誌D 情報・システム	6. 最初と最後の頁 459 ~ 470
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2019JDP7065	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 木下 涼、植野 真臣	4. 巻 J103-D
2. 論文標題 深層学習によるテスト理論 : Item Deep Response Theory	5. 発行年 2020年
3. 雑誌名 電子情報通信学会論文誌D 情報・システム	6. 最初と最後の頁 314 ~ 329
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2019JDP7068	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 菅原 聖太、植野 真臣	4. 巻 J103-D
2. 論文標題 Augmented Naive Bayes制約をもつベイジアンネットワーク分類器の厳密学習	5. 発行年 2020年
3. 雑誌名 電子情報通信学会論文誌D 情報・システム	6. 最初と最後の頁 301 ~ 313
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2019JDP7011	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 青見 樹、菅原 聖太、植野 真臣	4. 巻 J103-D
2. 論文標題 アンサンブル学習によるモデル平均ベイジアンネットワーク分類器	5. 発行年 2020年
3. 雑誌名 電子情報通信学会論文誌D 情報・システム	6. 最初と最後の頁 183 ~ 193
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2019JDP7059	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 宇都 雅輝	4. 巻 37
2. 論文標題 テスト理論と人工知能に基づくパフォーマンス評価の新技术	5. 発行年 2020年
3. 雑誌名 教育システム情報学会誌	6. 最初と最後の頁 8~18
掲載論文のDOI (デジタルオブジェクト識別子) 10.14926/jsise.37.8	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 八木 嵩大、宇都 雅輝	4. 巻 10
2. 論文標題 パフォーマンス評価における多次元項目反応モデル	5. 発行年 2019年
3. 雑誌名 電子情報通信学会論文誌D 情報・システム	6. 最初と最後の頁 708-720
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2019JDP7018	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 宇都 雅輝	4. 巻 J102-D
2. 論文標題 論述式試験における評点データと文章情報を活用した項目反応トピックモデル	5. 発行年 2019年
3. 雑誌名 電子情報通信学会論文誌D 情報・システム	6. 最初と最後の頁 553~566
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2019JDP7007	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 本田 和雅、名取 和樹、菅原 聖太、磯崎 隆司、植野 真臣	4. 巻 J102-D
2. 論文標題 推移性を利用した大規模ベイジアンネットワーク構造学習	5. 発行年 2019年
3. 雑誌名 電子情報通信学会論文誌D 情報・システム	6. 最初と最後の頁 796~811
掲載論文のDOI (デジタルオブジェクト識別子) 10.14923/transinfj.2019JDP7033	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

〔学会発表〕 計14件（うち招待講演 0件 / うち国際学会 5件）

1. 発表者名 Masaki Uto
2. 発表標題 A Multidimensional Item Response Theory Model for Rubric-based Writing Assessment
3. 学会等名 International Conference on Artificial Intelligence in Education (国際学会)
4. 発表年 2021年

1. 発表者名 Itsuki Aomi, Emiko Tsutsumi, Masaki Uto, Maomi Ueno
2. 発表標題 Integration of Automated Essay Scoring Models using Item Response Theory
3. 学会等名 International Conference on Artificial Intelligence in Education (国際学会)
4. 発表年 2021年

1. 発表者名 青見樹, 堤瑛美子, 宇都雅輝, 植野真臣
2. 発表標題 項目反応理論を用いた自動採点モデルの統合手法
3. 学会等名 第35回人工知能学会全国大会
4. 発表年 2021年

1. 発表者名 Masaki Uto, Yikuan Xie, Maomi Ueno
2. 発表標題 Neural Automated Essay Scoring Incorporating Handcrafted Features
3. 学会等名 International Conference on Computational Linguistics (国際学会)
4. 発表年 2020年

1. 発表者名 宇都雅輝・植野真臣
2. 発表標題 ルーブリック評価のための多相項目反応モデル
3. 学会等名 行動計量学会第48回大会
4. 発表年 2020年

1. 発表者名 新田森・宇都雅輝
2. 発表標題 ルーブリック評価における多次元4相型項目反応モデルの提案
3. 学会等名 人工知能学会 第91回 先進的学習科学と工学研究会
4. 発表年 2021年

1. 発表者名 Maomi Ueno, Yoshimitsu Miyazawa
2. 発表標題 Uniform adaptive testing using maximum clique algorithm
3. 学会等名 International Conference on Artificial Intelligence in Education (AIED) (国際学会)
4. 発表年 2019年

1. 発表者名 Masaki Uto
2. 発表標題 Rater-effect IRT model integrating supervised LDA for accurate measurement of essay writing ability.
3. 学会等名 International Conference on Artificial Intelligence in Education (AIED) (国際学会)
4. 発表年 2019年

1. 発表者名 岡野将士, 宇都雅輝
2. 発表標題 評価者バイアスに頑健な小論文自動採点手法
3. 学会等名 第88回 人工知能学会 先進的学習科学と工学研究会
4. 発表年 2020年

1. 発表者名 堤瑛美子, 宇都雅輝, 植野真臣
2. 発表標題 時系列学習データを用いた隠れマルコフIRTによる高精度パフォーマンス予測
3. 学会等名 第47回日本行動計量学会
4. 発表年 2019年

1. 発表者名 宇都雅輝
2. 発表標題 項目反応理論と機械学習技術を用いた小論文評価手法
3. 学会等名 日本テスト学会第17回大会
4. 発表年 2019年

1. 発表者名 堤瑛美子, 塩野谷周平, 宇都雅輝, 植野真臣
2. 発表標題 Bayesian Knowledge Tracingの一般化としての隠れマルコフIRTモデル
3. 学会等名 第33回人工知能学会全国大会
4. 発表年 2019年

1. 発表者名 宮澤芳光, 植野真臣
2. 発表標題 最適化問題を用いた等質適応型テスト
3. 学会等名 日本行動計量学会第47回大会
4. 発表年 2019年

1. 発表者名 Yoshimitsu Miyazawa, Maomi Ueno
2. 発表標題 Computerized Adaptive Testing Method using Integer Programming to Minimize Item Exposure
3. 学会等名 人工知能学会全国大会(第33回)
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	宇都 雅輝 (Uto Masaki) (10732571)	電気通信大学・大学院情報理工学研究科・准教授 (12612)	
研究分担者	宮澤 芳光 (Miyazawa Yoshimitsu) (70726166)	独立行政法人大学入試センター・研究開発部・助教 (82616)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------