

人間の技能やモノの品質をAIで高精度に評価する

宇都 研究室



宇都 雅輝
Masaki UTO

2020年度に始まる大学入学共通テストでは、鳴り物入りだった「記述式問題」の導入が見送られました。思考力や判断力、表現力を問う記述式の重要性は認識されていたものの、現体制では採点に時間がかかることや、公正な採点が難しいといった課題が見送りの理由でした。しかし、長期的にみれば、記述式のテストは今後多くの分野の試験に導入されていくのではないのでしょうか。

AIが採点

そのための研究は着実に進んでいます。宇都雅輝助教は、統計学や確率論、機械学習などを基礎にした人工知能(AI)技術を使って、多様な試験において、採点者によらない厳正な採点手法を提案したり、将来を見ずして人が介入せずすべて自動で採点する手法を考案したりしています。

宇都助教の最近の主な研究は、ショッピングサイトやグルメサイトなどで、ユーザが商品やお店に対して星の数などでレーティング(評価)した「レーティングデータ」を解析する手法の開発や、自動採点などに向けた統計的な自然言語処理の新たなモデルの開発など

です。特に教育評価に関するデータを扱うことが多く、理論と応用の両面から研究しています。

採点者によるバラつきを補正

レーティングデータには、商品などを評価したレビューデータだけでなく、大学入試や講義の記述式問題やレポートの採点データなどのほか、入社試験の面接や実技試験といった試験データなども含まれます。レーティングデータはさまざまな場面で収集され、活用されていますが、実際にはそのデータはレビュー(評価者)の特性に強く依存し、対象を正確に反映しているとは言いがたいのが現状です。

そこで開発したのが、甘いまた

は厳しいといった、

人による採点のバラつきを補正し、標準化した適正なスコアを割り出すレーティングデータの新しい解析モデルです。教育や心理測定の基本理論である潜在変数モデルを用いた「項目反応理論」の拡張モデルで、現時点で世界最高精度のスコアを算出できることが分かっています。

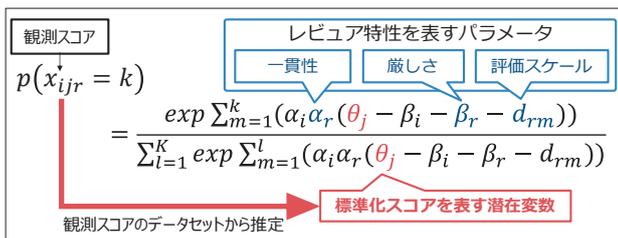
様々な試験で実用化

数理モデルの開発に加えて、公正な実技試験を実施するための試

験環境の設計や開発にも取り組んでいます。試験場にカメラを設置して試験風景を撮影し、管理サ

レビュー特性を考慮したスコアリングモデル

レビューの特性を考慮してスコアを標準化できる技術
項目反応理論と呼ばれる潜在変数モデルの拡張モデル



本研究室の技術が世界最高精度を達成

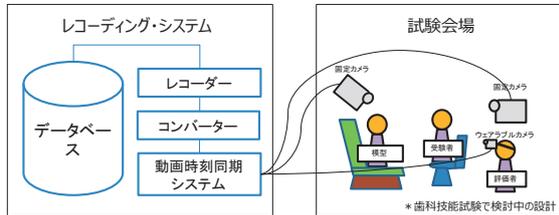
キーワード

データサイエンス、人工知能、機械学習、統計科学、確率論、自然言語処理、テスト理論、教育評価、情報システム

所属	大学院情報理工学研究科 情報・ネットワーク工学専攻
メンバー	宇都 雅輝 准教授
所属学会	電子情報通信学会、人工知能学会、言語処理学会、日本行動計量学会、日本テスト学会、日本教育工学会、教育システム情報学会、米コンピュータ学会(ACM)、国際人工知能教育学会(IAIED)
研究室HP	https://sites.google.com/view/utomasaki/

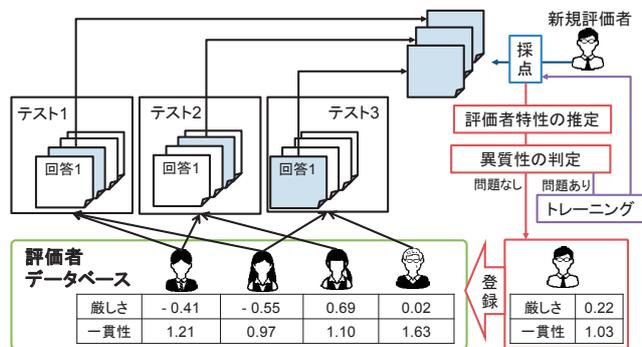
厳正な試験のための評価環境の設計・開発

試験の実施状況を記録・データベース化しておくことで、採点結果に疑義が生じた場合に第三者が確認したり、評価者トレーニングの素材として利用することが可能

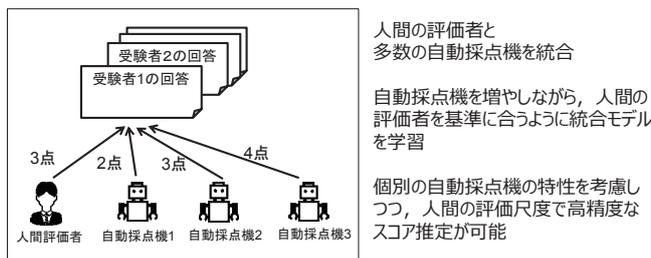


実技試験や語学試験，面接試験などで利用可能

評価者データベースとその活用



自動採点実用化を目指した今後の方針



人間と同様の採点精度を維持しつつ人間の負担を減らしていくことで
将来的に完全な自動採点を目指す

バにそのデータを蓄積しておくことで、事後の評価を可能にした。新たな採点者を育成するためのトレーニング教材として活用したりすることが出来ます。

入試や国家試験、入社試験などの試験は、その場合がその後のキャリアや人生を左右しかねないのにも関わらず、現状では「採点者ごとのバイアスを完全に排除することはできていない」と宇都助教は指摘します。宇都助教は、開発した解析モデルと試験環境シス

テムの実用化を進めており、これまでに医療系大学間共用試験(実技含む)や英検のスピーキングテスト、民間の教育関連会社の記述式問題、グループデスカッション形式の入社試験などに適用実績があります。

将来は、「不適切な採点者を発見してフィードバックしたり、その採点者に追加のトレーニングを課したり、またいずれは採点者をデータベース化して評価対象物ごとに最適な採点者を割り当てたり

することも可能になるだろう」と宇都助教は見通しを語ります。

完全自動採点にも道

もう一つの柱である統計的自然言語処理の研究では、大きな目標となるのが、コンピュータに完全に採点を任せる自動採点の実現です。自動採点技術は1990年代から研究され、2016年以降は深層学習(ディープラーニング)を使った高精度なモデルが普及し実用化が期待されています。

宇都助教は、既存のディープラーニング手法に項目反応理論を使った手法を融合させた、小論文や記述式問題向けの新しい自動採点モデルを開発しました。過去の採点データを学習させる際に、採点者によるバイアスを取り除いた精度の高いデータを用いることで、従来モデルよりも高い精度で自動採点が可能になりました。特定の言語に依存しないため、日本語だけでなくあらゆる言語の問題に適用できます。

今後、こうしたコンピュータによるさまざまな種類の自動採点機と人による採点を組み合わせることで、「記述式問題などの採点をより迅速かつ正確に行えるようになるれば、入試などへの導入の可能性がみえてくる」と宇都助教は考えています。さらに徐々に人による関与を減らしていけば、いずれは完全な自動採点が可能になるのかもしれない。

【取材・文】藤本信穂