

科学研究費助成事業 研究成果報告書

令和 2 年 5 月 19 日現在

機関番号：82616

研究種目：基盤研究(B) (一般)

研究期間：2017～2019

課題番号：17H01843

研究課題名(和文)人工知能を利用した短答式記述採点支援システムの開発

研究課題名(英文)AI-based Japanese Short-answer Scoring and Support System

研究代表者

石岡 恒憲 (Ishioka, Tsunenori)

独立行政法人大学入試センター・研究開発部・教授

研究者番号：80311166

交付決定額(研究期間全体)：(直接経費) 7,700,000円

研究成果の概要(和文)：短答式記述採点支援システムを実装した。採点エンジンの性能を評価するために、国立情報学研究所が主催するタスク競争型の国際研究集会NTCIR-13の質問応答のサブタスクに参加し、国内外11の研究機関が参加する中、横浜国立大学と並んでトップの成績を収めた。タスク課題は東大2次試験の世界史における500字論述試験の自動採点(過去5年分)である。研究成果は国際的なトップカンファレンスの一つであるWI 2007に採択された。その後も九州大学と連携しシステム性能の向上を目指した。技術的には、LSTMというストリーム型の言語モデルに学習の精度を上げるための工夫を創案し、JAWS 2019にその成果を発表した。

研究成果の学術的意義や社会的意義

記述式テストはいまままでの多肢選択テストに比べ、より正統あるいは真正で信頼できると広く考えられており、採点のための技術的な課題が克服されてさえいれば、これを使う潜在的な需要は計りしれない。本研究ではセンター試験など大学入試試験レベルの短答式記述試験の自動採点および人間による採点を支援する実用可能なシステムを試作・実装する。この研究は文部科学省の高大接続部会・最終答申にある「採点業務の効率・安定化のためにコンピュータの効果的な活用」や「人工知能の利用」に位置づけられる、重要かつ極めて緊急性の高い課題である。試作システムは最先端の技術を実装し、人工知能の活用の名に値するといっていよい。

研究成果の概要(英文)：We have developed a short answer scoring and support system. To evaluate the performance of the scoring engine, we participated in the task-answering international research meeting NTCIR-13 organized by the National Institute of Informatics, and won the top performer alongside Yokohama National University among the 11 participating research institutes around the world in the question-answering sub-task. The challenge of the competition is the automatic scoring of the 500-character essay examination in the world history of the University of Tokyo Second Examination for the past five years. The results were shown in WI 2007, one of the top international conferences. After that, we worked together with Kyushu University to improve system performance. We devised a stream-type language model called Long Short-term Memory (LSTM) to improve learning accuracy. The result was announced at the joint agent workshop & symposium (JAWS2019), which is well known in Japan.

研究分野：情報数理

キーワード：自動採点 自然言語処理 機械学習 深層学習 情報システム

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

文部科学省・高大接続部会の最終答申において、新テストに記述式テストが導入されることが決定し、採点業務の効率・安定化のためにコンピュータの効果的な活用や人工知能の利用が盛り込まれた。記述式テストには大別して「正解のないエッセイタイプ」のものと、「正解のある(したがって正解との意味の合致を判定する)短答式タイプ」のもの2つがあるが、いま新テストに求められている記述式は後者である。採点における技術的な困難から、質問の解答は数十文字から多くとも60文字程度であると予想されている。短答式テストはいままでの多肢選択テストに比べ、より正統あるいは真正(authentic)で信頼できると広く考えられており、採点のための技術的な課題が克服されてさえいけば、これを使う潜在的な需要は(センター試験に限らず)計りしれない。

エッセイタイプについてはこれまでも、特にアメリカにおいて多くのシステムが開発され実用に供されてきた。アメリカのビジネススクール入学のための共通テストであるGMATにおける作文(エッセイ)テストでは、1998年よりe-raterが、2006年よりはIntelliMetricが採点を行っている。他にも商用のシステムとして、PEG(Project Essay Grade)やIEA(Intelligent Essay Assessor)、CRASEなどがあり利用に供されている。我が国においても、我々のグループが日本語で初めての小論文自動採点システムJess(Ishioaka, 2006)を開発し、実用化の段階にある。成果については、朝日新聞夕刊の一面トップ(2005年2月)に大きく掲載されたほか、Yahoo!インターネットガイド(2006年6月)、韓国KBSテレビ(2007年2月)等、多くのマスコミで紹介された。しかし今ここで開発しようとしている短答式テストの自動採点については、その重要性は認められているものの技術的にさまざまな課題が未解決のままである。Vigilante(1999)は世界最大のテスト機関であるETSとニューヨーク大学とで、この分野における共同研究を行い、最初の報告をした。Leacock&Chodorow(2003)は、ETSが開発したc-raterの最新の仕様について報告している。Pulman & Sukkarieh(2005)は、情報抽出技術に隠れマルコフモデルなどの自然言語処理を用いて、システムが用意する正解文と同じ意味の文を幾つか自動生成する試みについて述べている。しかしその性能は、エッセイの自動評価採点システム(e-rater)に比べ、人間による専門家との一致率は10%以上も小さい。我が国においても国立情報学研究所の「ロボットは東大に入れるか」プロジェクトで、受験ロボットがセンター試験の選択肢の正誤を、教科書からの知識源との含意関係認識技術を用いて解くということを試みてきた。その結果わかってきたことは、完全な含意関係認識技術すなわち正しい意味理解は現時点では困難であり、いくつかの手法を組み合わせると半ばアドホックに解くことで、あたかも人間が解答したかのようにみせかけることがせいぜいであるという事実である。我々もいままでの科研「挑戦的萌芽研究; H23~25; 短答式記述テストにおける自動採点システムの開発」で自動採点に向けた要素技術の習得やシステム試作に努めてきたが、システムの限界もわかってきた。

2. 研究の目的

センター試験など大学入試試験レベルの短答式記述試験の自動採点および人間による採点を支援する実用可能なシステムを試作・実装することを目的とする。自然言語におけるテキスト間の同義や含意、推論を公的な試験に採用するほどの精度はこの数年では不可能であるという判断のもとに、採点は設問ごとに作題者が用意した「採点基準」に従いシステムがある程度(80%以上程度)の精度をもった採点計算(自動採点)を基本とし、その結果を人間が確認・修正できるものとする。

完全な自動採点システムとせず人間(採点者)が修正しうる余地を残すのは、現在の技術レベルでは正確な意味理解はまだ難しいことと、センター試験の後継となる新テストの記述には比較的自動採点が容易と思われた(教科書に書かれている事実との意味的照合で対応可能な)理科や社会の問題ではなく、国語に記述問題が採用されたことによる。

本研究は高大接続部会・最終答申にある「採点業務の効率・安定化のためにコンピュータの効果的な活用」や「人工知能の利用」に位置づけられる極めて緊急性の高い課題である。また記述テストの採点支援については現時点では文科省に答案のクラスタリングしか案がなく、ここで実装されるシステムはまさに人工知能の活用の名に値する新規性・独創性のある課題達成といえる。

短答式試験の完全なる自動評価採点支援システムは、完成した場合多くの教育関係者が実際の使用を欲するものである。そのモチベーションとしては第1に経済性が挙げられる。もし人間による評価採点の大半をコンピュータに置き換えることができれば、試験の評価に対するコストを大幅に削減することができる。アメリカのビジネススクールでの入学共通試験であるGMATにおける作文試験での適用の仕方と同様に、最終判定点の決定が人間側に委ねられてさえいけば、コンピュータ採点は一般には受け入れ易いと考えられる。2番目のモチベーションはより重要であるが、短答式の自動採点はシステムがよくできてさえいけば、即座のフィードバックを返すことができる。本システムでも採点基準への適合の程度をシステムが自動的に判定するから、初心者の採点者はこれを見ることで採点行為そのもののやり方を学習訓練することができる。これは個人教育的な側面を有しているともいえる。他のモチベーションとして、近年では説明責任が挙げられる。採点の論拠を示

すことの重要性はますます増大してきている。本システムはこれらの社会的要請へ応えるものといえる。

3. 研究の方法

大学入学試験の実際の記述問題や予備校の実施する試験問題を分析し、現時点で最も取り組みやすい比較的解答の短い問題で、かつ高度な判断による寄与の少ない理科や社会の問題に対してアプローチを行い、ある程度の成果を上げる。その後「国語」の採点に必要なより深い知的処理が必要な問題や暗黙の知識・推論を利用する技術について検討を行う。表層的な部分に加えて、模範解答との含意や同義関係認識を主な研究課題とする。入力が自然言語であるために自然言語処理とりわけ構文解析処理は必須であるが、他にも知識処理・推論、意図や比喻など認知に基づく意味理解などさまざまな種類の知的情報処理が必要である。

我々のアプローチは、

- (1) 設問ごとに作題者が用意した「採点基準」に従い、用意された解答がその採点基準の各項目に合致しているか否かを表層的な観点からシステムが自動判定し、
- (2) その自動判定に基づいた仮採点(チェック得点と称する)を与える他、システムはすでに専門家が採点した採点データ(教師データ)をもとに機械学習器で与えられる予測得点(システム得点と称する)の両方を与え、
- (3) 人間の採点者は、これらの情報を参考に、システムの提示する得点を認証することができる他、自らの判断でこれを修正・上書きできるものとする。ただしシステムの提示値は、ある程度の精度を有するものでないと、採点の手間を軽減するという点で有効とはいえ、80%以上の精度を一応の目標とする。機械学習に際しては、「採点基準」への表層的な一致に加え、「模範解答」と「(実際の)記述解答」との意味的近似性など、深層的な変量も考慮するものとする。

研究課題に対する大よその分担は以下の通り：

石岡(研究代表者): 総括、自然言語処理、システム設計、機械学習
峯(研究分担者): 自然言語処理、システム構築、評価手法の妥当性評価
宮澤(研究分担者): システム実装、JSP プログラミング
橋本(研究分担者): システム評価

本システムは Web アプリケーションとして提供するが、全体設計や Web システムの構築については石岡が担当する。評価手法の妥当性評価を峯が行った。ユーザインターフェイスや画面設計周辺については宮澤が担当した。橋本がシステムを評価した。

4. 研究成果

【平成 29 年度】

国立情報学研究所が主催するタスク競争型の国際研究集会 NTCIR-13 のうち質問応答のサブタスク(QALab-3 タスク)に我々の研究グループ(大学入試センター・九州大学合同チーム)で参加した。今回のコンペの課題は東大 2 次試験の世界史(過去 5 年分)における 500 字論述試験の解答自動作成とその自動採点である。このサブタスクには国立高雄応用科技大学、横浜国立大学、淡江大学、名古屋大学、京都大学、早稲田大学、カーネギーメロン大学(2 チーム)、慶應義塾大学、DG Lab に加え我々の計 11 チームが参加した。自動採点の評価において、我々は横浜国立大学と並んでトップの成績を収めることができた。日本語記述の自動採点の分野において、フロントランナーであることが改めて証明された。また本システムについてはトップカンファレンスの一つである Web Intelligence(ライブチヒ)に採択され、発表を行った。

【平成 30 年度】

九州大学システム情報科学院との連携をより強め、優秀な九州大学の学生とも協力し、システム性能の向上を目指した。人工知能隆盛の現在にあって、自然言語処理分野でよい成果を上げている LSTM というストリーム型の言語モデルを使って採点を行う方法が有望であることがわかりつつあり、我々もその研究に追隨した。成果については手書き文字認識の第一人者である中川正樹教授(農工大)の招きによりカナダ・モントリオールで開催された国際会議 ICPRAI2018 で口頭発表した。また 9 月には日本テスト学会で、エッセイおよび論述解答試験採点に関する最近の話題と自動採点にむけた現在の技術水準について講演を行った。

【令和元年度】

自然言語処理分野でよい成果を上げている LSTM というストリーム型の言語モデルを使う方法に、学習の精度を上げるために教師データをかさ上げすべく品詞別ノイズ付与の方法を創案し、その性能を評価した。成果については国内では著名な合同エージェントワークショップ&シンポジウム 2019(JAWS2019)に採択され、口頭発表した。またひつじ書房より「外国語教育研究における自動採点の現状と課題」と題する出版計画があり、その本の総括として最終章の「自動採点研究のこれから」と題する章を執筆した。現在校正段階にあり 2020 年中には刊行される予定である。

5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 6件/うち国際共著 5件/うちオープンアクセス 0件）

1. 著者名 Tsunenori Ishioka and Masayuki Kameda	4. 巻 ICPRAI 2018
2. 論文標題 AI-based Automated Japanese Short-answer Scoring and Support System	5. 発行年 2018年
3. 雑誌名 International Conference on Pattern Recognition and Artificial Intelligence	6. 最初と最後の頁 138-143
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 該当する
1. 著者名 山口 晃平, 峯 恒憲	4. 巻 Vol. J101-D, No.9
2. 論文標題 地域問題レポートに対する担当課推定手法の提案	5. 発行年 2018年
3. 雑誌名 電子情報通信学会論文誌：特集号「ソフトウェアエージェントとその応用」	6. 最初と最後の頁 1334-1342
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 Jihed Makhoulouf and Tsunenori Mine	4. 巻 EDM 2018
2. 論文標題 Predicting if students will pursue a STEM career using School-Aggregated Data from their usage of an Intelligent Tutoring System	5. 発行年 2018年
3. 雑誌名 EDM 2018 poster	6. 最初と最後の頁 533-536
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 Tsunenori Mine and Sachio Hirokawa and Takahiko Suzuki	4. 巻 IMCOM2019
2. 論文標題 Does Crime Activity Report Reveal Regional Characteristics?	5. 発行年 2019年
3. 雑誌名 The 13th International Conference on Ubiquitous Information Management and Communication	6. 最初と最後の頁 Session 8:3
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 Ishioka Tsunenori, Kameda Masayuki	4. 巻 WI 2017
2. 論文標題 Overwritable automated japanese short-answer scoring and support system	5. 発行年 2017年
3. 雑誌名 The IEEE/WIC/ACM International Conference on Web Intelligence,	6. 最初と最後の頁 50-56
掲載論文のDOI (デジタルオブジェクト識別子) 10.1145/3106426.3106513	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 該当する

1. 著者名 Tsunenori Ishioka, Kohei Yamaguchi, Tsuneori Mine	4. 巻 NTCIR 13
2. 論文標題 Rubric-based Automated Japanese Short-answer Scoring and Support System Applied to QALab-3	5. 発行年 2017年
3. 雑誌名 Proceedings of the 13th NTCIR Confereof Information Access Technologiesnce on Evaluation	6. 最初と最後の頁 152-158
掲載論文のDOI (デジタルオブジェクト識別子) なし	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 該当する

1. 著者名 Yao Lin, Kohei Yamaguchi, Tsunenori Mine, Sachio Hirokawa	4. 巻 SCDM 2018
2. 論文標題 Is SVM+FS better to satisfy Decision by Majority ?	5. 発行年 2018年
3. 雑誌名 The 3rd International Conference on Soft Computing and Data Mining 2018	6. 最初と最後の頁 261-271
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-319-72550-5_26	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 該当する

1. 著者名 Sachio Hirokawa, Takahiko Suzuki, Tsunenori Mine	4. 巻 WI 2017
2. 論文標題 Machine Learning is better than Human to satisfy Decision by Majority	5. 発行年 2017年
3. 雑誌名 The IEEE/WIC/ACM International Conference on Web Intelligence	6. 最初と最後の頁 694-701
掲載論文のDOI (デジタルオブジェクト識別子) なし	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 該当する

〔学会発表〕 計8件（うち招待講演 2件 / うち国際学会 2件）

1. 発表者名 加藤博之, 石岡恒憲, 峯恒憲
2. 発表標題 短答式試験における自動採点のための品詞別ノイズ付与手法の提案
3. 学会等名 合同エージェントワークショップ&シンポジウム2019 (JAWS2019)
4. 発表年 2019年

1. 発表者名 石岡恒憲, 劉東岳, 峯恒憲
2. 発表標題 エッセイおよび論述解答試験採点に関する最近の話題と自動採点にむけた現在の技術水準
3. 学会等名 日本テスト学会第16回大会
4. 発表年 2018年

1. 発表者名 Jihed Makhoulouf and Tsunenori Mine
2. 発表標題 Investigating How School-Aggregated Data Can Influence in Predict
3. 学会等名 EDM 2018 Workshop on Scientific Findings from the ASSISTments Longitudinal Data (国際学会)
4. 発表年 2018年

1. 発表者名 峯 恒憲, 廣川 佐千男, 鈴木 孝彦
2. 発表標題 警察の犯罪レポートメールに基づく地域の識別
3. 学会等名 合同エージェントワークショップ&シンポジウム2018
4. 発表年 2018年

1. 発表者名 Jihed Makhoulf and Tsunenori Mine
2. 発表標題 Investigating Reading Behaviors within Student Reading Sessions to Predict their Performance
3. 学会等名 International Workshop on Predicting Performance Based on the Analysis of Reading Behavior:A Data Challenge 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 石岡 恒憲
2. 発表標題 人工知能を利用した記述採点支援システム
3. 学会等名 高校教科「情報」シンポジウム2017秋, 情報処理学会 (招待講演)
4. 発表年 2017年

1. 発表者名 石岡 恒憲
2. 発表標題 記述解答採点におけるAI応用の現状と将来
3. 学会等名 ICT CONNECT 21 (招待講演)
4. 発表年 2017年

1. 発表者名 山口 晃平, 峯 恒憲
2. 発表標題 地域問題レポートに対する担当課の自動推定
3. 学会等名 合同エージェントワークショップ&シンポジウム 2017(JAWS2017)
4. 発表年 2017年

〔図書〕 計1件

1. 著者名 石岡恒憲 (石井雄隆・近藤悠介 編)	4. 発行年 2020年
2. 出版社 ひつじ書房	5. 総ページ数
3. 書名 自動採点研究のこれから (分担執筆) 「外国語教育研究における自動採点の現状と課題」	

〔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	峯 恒憲 (Mine Tsunenori) (30243851)	九州大学・システム情報科学研究院・准教授 (17102)	
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)	
研究 分担者	橋本 貴充 (Hashimoto Takamitsu) (20399489)	独立行政法人大学入試センター・研究開発部・准教授 (82616)	