

令和 5 年 6 月 20 日現在

機関番号：54401

研究種目：基盤研究(C) (一般)

研究期間：2018～2022

課題番号：18K02886

研究課題名(和文) 筆記試験答案を対象としたデータ分析のための前処理の検討と支援システムの開発

研究課題名(英文) The development of the system for assisting teachers in the pre-processing of written answers for analysis

研究代表者

和田 健 (Wada, Takeshi)

大阪公立大学工業高等専門学校・その他部局等・准教授

研究者番号：00469587

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：本研究は、筆記試験に対する答案分析の「前処理」に焦点を当て、その実務作業を支援するシステムの研究開発に取り組んだ。前処理は、紙媒体の答案から必要情報を抽出し、そこから定量分析が可能な構造化データを生成する処理を指す。前処理の手順と内容について調査と分析を行い、それに基づき、使いやすさ、拡張性、汎用性を備えたWindowsアプリケーションとPythonスクリプトから構成されるシステムを開発した。そして、実際の試験答案を使用した性能評価を行ない、短時間でミスが減らして前処理を可能にすることを実証的に確認した。

研究成果の学術的意義や社会的意義

この研究は、教育ビッグデータとしての筆記試験答案の分析のための「前処理」を自動化・効率化する作業支援システムを提案・開発したものである。学術的には、従来困難とされてきた筆記解答の大量データ処理を効率的に行うことで、学習者の理解度や学習行動の詳細な把握が可能となる。これにより、教育方法の改善や個々の学習支援の具体的な方策を見つけ出す新たな研究フィールドが開かれる可能性がある。社会的には、教育現場における負担軽減や教育の質向上に寄与し、個々の学習者への最適な指導が可能となり、教育格差の解消にも繋がると考えられる。

研究成果の概要(英文)：In this research, we focused on "pre-processing" in analyzing written examination answers and developed a system to facilitate this process. Pre-processing involves extracting critical information from paper-based answer sheets and converting it into data files for computer software analysis. Our first steps included surveying and analyzing the tasks and procedures necessary for answer sheet pre-processing. We then deliberated on a system allowing teachers to accomplish this task efficiently via computers. Subsequently, we crafted, implemented, and introduced a system comprising three user-friendly Windows applications and an extensible, versatile Python script. This system's functionality and performance in supporting actual exam paper pre-processing were evaluated. The outcome verified that pre-processing with our system can be achieved faster than manual work, error-free.

研究分野：システム工学

キーワード：ラーニングアナリティクス 答案分析 データ分析前処理 教育工学

## 1. 研究開始当初の背景

中等・高等教育機関では、学習者の学習到達度の測定を目的として「定期試験」が実施されている。また、日常的な学習の促進や知識の定着を目的として授業のなかでは「小テスト」も実施されている。これら試験の答えは、教員により採点され、主に成績評価の資料として用いられている。また、個々の学習者の苦手や弱点を把握した効果的な個別指導や、教科教育法の評価や改善に向けた資料としても有効に活用されている。

しかし、試験答案が有する情報のうち一般に定量的な分析の対象となっているのは、「得点」として縮約された 1 次元の数値データに限られ、その他の情報については採点者の定性的な分析にとどまることが多い。例えば、2 名の学習者の得点と同じであれば、両者は同程度の学力を持つという分析・判断がされる。しかし、得点と同じであっても、答案における正誤箇所には異なるはずであり、その観点から得点だけに依存しない、より深い答案分析による学力評価が求められる。しかし、現実的には以下の理由からそのような分析を行なうことは難しい。

- (1) 答案が紙媒体であり、解答が記述式（非マークシート形式）で記入されていること
- (2) 試験の実施頻度、問題数、学生数の乗算で情報量が増大するビッグデータであること
- (3) 分析目的に応じて、問題とその解答に対して様々な属性を手動で与える必要があること  
問題の属性例：基礎、応用、作図、計算、証明、出題予告した問題、教科書掲載問題  
解答の属性例：正解、不正解、部分正解、空欄解答、計算誤り、有効数字誤り

試験答案は得点以外にも潜在的に有意義な情報を有するデータソースと言えるが、その情報を電子化・抽出・集約・加工して定量分析が可能な状態（＝構造化データ）に整える「前処理」が非常に煩雑であるという特性を持つ。事実、効果的な教育指導の実現のためにビッグデータの利活用が模索されているが、筆記試験答案を教育ビッグデータとして捉えて、その分析や利活用を試みている事例はほとんど見られない。得点という評価指標には表れない情報の定量分析を通して教育指導上の新たな知見や洞察を得られる可能性は高く、そのボトルネックとなっている前処理について明らかにし、それを支援するシステムについて研究開発することが求められている。

## 2. 研究の目的

本研究の目的は、中等教育機関や高等教育機関で実施される記述式の試験答案のデータ分析に必要な前処理プロセスを明らかにし、そのプロセスを省力化・効率化するための実用的な支援システムを開発することにある。

具体的には、紙媒体の記述式の試験答案を対象に、必要情報を抽出して統計解析ソフトで取り扱い可能な構造化データを生成するまでの「前処理」に着目し、そこに求められる手順や内容を明らかにする。そして、その前処理にかかる作業を、補助あるいは省力化・自動化するための手法を検討し、それを実装した作業支援システム（ソフトウェア）を開発し、その効果について評価する。さらに、実際の筆記試験答案を用いて、どのようなラーニングアナリティクスが可能であるか検討する。

## 3. 研究の方法

本研究は、以下の 3 つのサブテーマに分けて取り組んだ。

第一に、研究者が所属機関で担当している授業で実施する小テストや定期試験の答案を対象に、汎用の統計解析ソフトを用いてデータ分析するために必要な前処理のプロセスについて詳細に調査した。具体的には、学力評価や学習支援などの観点から様々な分析シナリオを検討し、そこで共通する前処理プロセスを実証的に明らかにした。その結果から紙媒体の答案から統計解析ソフトに入力可能な構造化データを生成するための手順について定義した。

第二に、前処理に係る作業を補助・自動化する支援システムの設計と実装に取り組んだ。このシステムは、前処理の各プロセスに必要な機能を提供する複数の GUI アプリケーションの組み合わせとして開発した。具体的には、スキャンした答案画像の傾きや位置を確認・調整するツール、答案画像を解答欄単位にトリミング/分解するツール、解答欄単位に分解された画像に対して分析のための属性を付与する作業を補助するツールについて開発と継続的な改良に取り組んだ。また、支援システムを使用することによる時間短縮の効果を評価するための実験を行った。

第三に、筆記試験答案を対象とした定量的なデータ分析を通じて、どのようなラーニングアナリティクスが可能であるかについて検討した。そして、支援システムと組み合わせ使用可能な分析シナリオのテンプレート（Python スクリプト）を作成した。

#### 4. 研究成果

##### (1) 筆記試験答案の定量的データ分析に必要な前処理についての分析

紙媒体の記述式の筆記試験答案から、データ分析に必要な情報を抽出した構造化データを生成するために必要な共通プロセスについて明らかにした。具体的には、答案のスキャン(画像データ化)、画像補正(スキャン時に発生する位置や傾きのずれの修正)、画像と学習者IDの紐づけ、答案全体画像を解答欄単位の画像に分解・トリミング、解答欄単位の画像(=学習者の解答)に対する属性情報の付与、各試験問題に対する属性情報の付与、統計解析ソフトに入力可能な構造化データの生成(データの結合と整形)という作業が前処理のプロセスとして必要となることを明らかにした。

##### (2) 前処理のための作業支援システムの開発

前処理のための各作業を効率化・自動化するためのツール群を設計実装し、それらを組み合わせて前処理全体について支援が可能なシステムを開発した。

第一に、図1に示す画像補正/リネームツールを開発した。このツールはスキャンした答案画像と学習者IDをファイル名として紐づける作業と、連続スキャン時に生じる画像の位置や傾きを確認・調整する作業を補助する機能を提供する。

第二に、図2に示す一括トリミングツールを開発した。このツールはマウス操作により解答欄位置を指定し、指定フォルダ内に存在する複数の答案画像に対して画像トリミング処理を一括適用し、解答欄単位の画像を生成する機能を提供する。

第三に、図3に示すように、解答欄単位の画像に対して、任意の属性情報を付与するツールを開発した。このツールにより個々の学習者解答に対して「正答」や「誤答」のほか、「スペルミス」や「空欄」のような分析のための属性情報を効率的に与えることが可能となる。

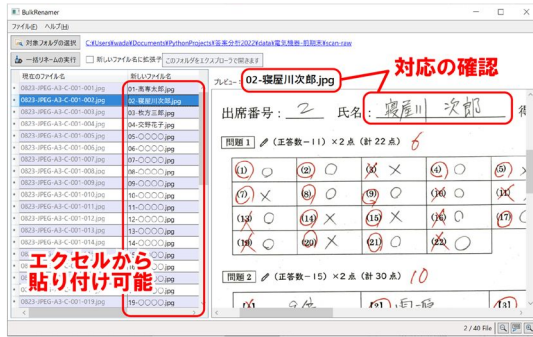


図1 画像補正/リネームツール

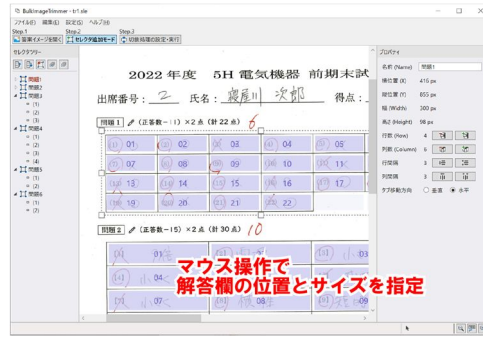


図2 一括トリミングツール



図3 属性情報付与ツール

第四に、以上の各ツールからの出力データや、分析者が作成する問題についての属性情報(例えば「基礎問題」や「応用問題」「計算問題」)などを結合・整形して、汎用の統計解析ソフトに入力可能なCSVやJSON形式の構造化データを出力するPythonスクリプトを作成した。

また、これらのツールを組み合わせさせたシステムを使用した場合の時間短縮の効果を測る評価実験を実施した。その結果、手作業で前処理に取り組んだ場合と比較して30~60%の時間で前処理が可能となることが実証的に確認された。

##### (3) ラーニングアナリティクスのための答案分析シナリオの作成

汎用的な答案分析シナリオを作成し、その分析作業を自動化するPythonスクリプトを作成した。具体的には、試験の各問題に「科目達成目標との対応」を属性として与え、達成目標別に試験の正答率を評価する分析シナリオや、ベイズ推定を利用して各問題の正誤から学習者の理解度を推定する分析シナリオを作成した。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 和田 健, 早川 潔, 谷野 圭亮	4. 巻 56
2. 論文標題 筆記試験答案分析のための前処理支援システムの開発	5. 発行年 2022年
3. 雑誌名 大阪公立大学工業高等専門学校 研究紀要	6. 最初と最後の頁 5-10
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 オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計4件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 筆記試験答案分析支援システムを活用した答案分析の事例報告
2. 発表標題 和田 健, 早川 潔, 谷野 圭亮
3. 学会等名 日本高専学会第28回年会講演会
4. 発表年 2022年

1. 発表者名 和田 健, 早川 潔, 谷野 圭亮
2. 発表標題 画像分類技術を利用した答案採点支援システムの研究開発
3. 学会等名 日本高専学会第27回年会講演会
4. 発表年 2021年

1. 発表者名 和田 健, 早川 潔, 谷野 圭亮
2. 発表標題 筆記試験答案を対象としたデータ分析シナリオの検討と考察
3. 学会等名 日本高専学会第26回年会講演会
4. 発表年 2020年

1. 発表者名 和田 健, 早川 潔, 谷野 圭亮
2. 発表標題 記述式筆記試験答案分析のための作業支援ツールの研究開発
3. 学会等名 日本高専学会第25回年会講演会
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	早川 潔  (Hayakawa Kiyoshi)  (20325575)	大阪公立大学工業高等専門学校・その他部局等・教授   (54401)	
研究分担者	谷野 圭亮  (Tanino Keisuke)  (70778589)	大阪公立大学工業高等専門学校・その他部局等・講師   (54401)	

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
---------	---------