

筆記試験答案分析のための前処理支援システムの開発

和田健*, 早川潔*, 谷野圭亮**

The development of the system for assisting teachers
in the pre-processing of written answers for analysis

Takeshi WADA*, Kiyoshi HAYAKAWA* and Keisuke TANINO**

ABSTRACT

In this study, researchers focused on "pre-processing" for analyzing written examination answers, and conducted research and development of a system to support it. Pre-processing refers to the process of extracting necessary information from paper-based answer sheets and creating data files that can be read by analysis software running on a computer. First, we surveyed and analyzed the tasks and procedures required for the preprocessing of answer sheets for analysis. Second, we discussed a system that would allow teachers to perform this task efficiently with computers. Then, we designed, implemented, and released a system whose elements were three Windows applications with easy-to-use interfaces and one Python script with extensibility and versatility. We evaluated the system's functionality and performance to support preprocessing of actual exam papers. As a result, it was confirmed that the pre-processing using the system can be performed in a shorter time than manual work without any errors.

Key Words: learning analytics, answers analysis of examination, pre-processing, paper-based answer sheets

1. はじめに

近年、教育や学習に関する様々な情報を「教育ビッグデータ」として分析・活用するラーニング・アナリティクス (LA : Learning Analytics) が注目を集めている。LA は、データという根拠 (エビデンス) に基づく教育改善や学習支援を目的としている^[1]。LA の代表的な事例として、e ラーニングシステムと連携した取り組みを挙げることができる。そこでは、教材に対するアクセスログ分析 (閲覧数、滞在時間、移動元/移動先) に基づく学習コンテンツの改良、学習行動履歴を可視化した学習ポートフォリオの提供、将来成績や履修途中放棄の予測などが試みられている^[2]。

ところで、教育ビッグデータのなかでも、各種試験を通じて学習者によって作成される試験答案は、到達度評価や授業改善に対して極めて有用な情報を含んだデータソースと言える。CBT (Computer Based Testing) やマークシート方式であれば試験解答は CSV 等の構造化データとして得られるため、統計解析ソフトを利用して直ちに定量分析ができる。

しかし、高専をはじめ中学や高校で実施される試験では、解答を文字で紙に記述する方式 (以下、筆記解答方式と称する) が広く採用されており、これらは分析データとしてコンピュータで直接的に扱うことはできない。紙媒体の答案をコンピュータで扱うためには、スキャン等による電子化、データの抽出、さらに分析の目的にあわせた変換、加工、集約といった「前処理」が必要となる。しかし、これら前処理を教員が手作業で行なうことは、時間的にも労力的にも大きな困難がともなう。また、答案は科目によりフォーマットが大きく異なるため、その前処理を自動化することは技術的に困難がある。実際に、筆者らが調査した限りでは、前処理を自動化できるようなツールは有償・無償を問わずに国内には見当たらない。

本研究では、このような筆記解答方式の答案分析に必要な「前処理」に対して、その内容と手順を明らかにし、その省力化と効率化を目的とした作業支援システムの提案と開発に取り組んだ。そして、前処理の各作業を補助するためのツールとスクリプトを作成し、その利用を前提とした作業フローについて提案した。さらに、本校で実施された定期試験の答案を対象として前処理を実施し、作業支援システムの機能と性能について評価した。

2022年8月31日 受理

* 総合工学システム学科 知能情報コース
(Dept. of Technological Systems : Intelligent Informatics Course)

** 一般科目 (General Education)

2. 答案分析の目的と期待できる効果

本研究において対象とする答案は、中間試験や期末試験といった定期考査で実施する筆記解答方式の答案となる。答案分析の目的を明確にするために、まずは、定期考査の実施目的や位置づけについて整理しておく。これらは教科や教員によって違いはあるが、概ね次のようなものと考えることができる。

- ① 学生個人もしくはクラス・学年について、学習到達目標に対する達成度を評価・判断するため。
- ② 進級／配属／推薦などの基準に利用可能な学力数値指標を、公平かつ客観的であると言える方法により得るため。
- ③ 教科指導法の効果や成否を評価・判断するため。
- ④ 学生が、自身の学力や傾向、課題を認識・把握し、学習の方法や量を見直すため。
- ⑤ 学生に対して（試験勉強という名目による）自己学習や復習を強く促すため。

上記のような目的で実施された試験は、教員によって答案のマル付け（＝設問単位の正誤評価）がされる。さらに（主として②の観点から）配点という重みをつけて正答数を合計した 100 点満点の得点が答案に記載され、学生に返却される。

ここで得点（総得点）とは、答案の内容を 1 次元に縮約した数値データになる。扱いやすく単純明快な数値データであるが、それゆえに（答案が有する様々な情報を置き去りに）独り歩きしてしまうことも多い。例えば、学生の意識は「合格基準である 60 点を超えたかどうか」や「友人と比較して得点が高いかどうか（勝ち負け）」に向けられ、答案返却の本来の目的である④に意識が向かないことがしばしば起こる。得点の高低は、実際問題として配点の加減によるところも大きく、本質的には「62 点であれば達成目標のある段階に到達しており、58 点であれば到達していない」のような判断ができる情報とはならない。また、2 人の学生が同じ得点であっても、その内訳（誤答した問題や誤答の内容）は一般には異なるはずであり、得点と同じことが、すなわち同じ学力を意味することにはならない。例えば、「私は、昨日、英語の勉強をしました」を英訳せよという問題に対して 2 人が同じく誤答であっても、その内容に「I study English yesterday」と「Me yesterday English study.」という違いがあれば、その学力が同じであるとは評価できない。

同時に教員側でも、得点の要約統計量（平均値、中央値、標準偏差、最高値、最低値など）や、その

変化に基づき①や③を判断・考察してしまうことも多い。例えば、学力補充指導の対象者の選定、授業速度の調整、演習の量や難易度の加減などである。しかし、先述のように「答案」から「得点」に縮約されるプロセスでは、多くの情報の欠落があり、得点だけに基づいて①や③をすることは十分とは言えない。

無論、得点化することや、それを利用することは②の観点から必要であり、否定されるものではない。現状においても、学生に対しては、答案返却時の解説や、間違えた問題の解き直しなどを通じて、④を強く意識させることはできる。しかし、それに加え、答案分析に基づく知見や洞察を組み合わせることで①③④に対して、より効果的かつ科学的なアプローチが可能になると考えられる。これらが、本研究において考える答案分析の目的ならびに期待できる効果である。

3. 筆記解答方式答案の分析に必要な前処理の検討

筆記解答方式の答案全数を対象とするような定量分析に必要な前処理について検討した。定量分析とは数値などの定量データを扱った分析の総称であるが、統計解析ソフトや表計算ソフトを利用して行なうためには電子化された構造化データ（つまり CSV ファイルなど）が必要となる。本研究では、答案の定量分析に対して汎用的かつ柔軟に利用できる構造化データとして「解答属性データ」と「設問属性データ」を考えた。そして、紙媒体の答案を入力として、これらのデータを CSV ファイルとして出力するための一連の処理を「前処理」と定めた。

解答属性データは、各答案の個々の解答記述に対して各種属性の有を「1」、無を「0」と表現した数値データとなる。ここで属性とは、分析者が分析目的ならびに各設問の特性に基づき任意に与えるものとなる。例えば「正答」「誤答」のほか、目的に応じて「空欄（解答記述なし）」「ケアレスミス」「誤字・スペルミス」「文字判読困難」「有効数字のミス」「途中式省略」「解答指示違反」のようなものになる。解答属性データの一例を表 1 に示す。

表の縦方向の見出しは学生を識別する情報となる。また、横方向の見出しは問題番号と属性を表す情報となる。例えば、表のなかで「01-高専太郎」の「問問題 1-正答」が「1」となっているが、これは、「高専太郎」の答案において問題 1 の解答記述が「正答」という属性を持つことを意味している。このデータを集計することで、各設問の正答率（誤答率）や空

欄率のような定量評価が可能となる。また、各学生について、ケアレスミスの数量、判読困難な解答記述の数量などを定量的に把握することもできる。さらに、多肢選択式の記号解答問題において、解答そのもの（つまり「ア」「イ」「ウ」など）を属性とすれば、その解答割合について定量的な情報を得ることもできる。

表1 解答属性データの例

	問題1	問題1	問題1	問題1	問題2	問題2	問題2	問題2	問題3	問題3	...	問題25	
	正答	誤答	空欄	有効数...	正答	誤答	空欄	誤字	正答	誤答	空欄	定数使用	
01-高専太郎	1	0	0	0	0	1	0	0	1	0	0	...	0
02-寝屋川次郎	0	1	1	0	0	1	1	0	1	0	0	...	1
03-枚方三郎	1	0	0	0	1	0	0	0	0	1	0	...	0
⋮													
40-門真末吉	1	0	0	0	0	1	1	0	0	1	1	...	0

一方、設問属性データは、設問そのものに対する各種属性の各種属性の有無を表した数値データとなる。分析目的に応じて「基本問題」「応用問題」「計算問題」「作図問題」「証明問題」「多肢選択式解答」「〇×選択式解答」「出題予告した問題」「過去に出題した問題」「ループリック観点1(学習到達目標1)」「ループリック観点2」などが属性となる。設問属性データの一例を表2に示す。

表2 設問属性データの例

	基本問題	応用問題	多肢選択式問題	記述式問題	作図問題	ループリック観点1	ループリック観点2	ループリック観点3	ループリック観点4	...	出題予告
問題1	1	0	0	1	0	1	0	0	0	...	1
問題2	1	0	0	1	0	0	1	0	0	...	1
問題3	1	0	1	0	0	1	0	0	0	...	0
⋮											
問題25	0	1	0	0	1	0	0	1	0	...	0

解答属性データと設問属性データの2つをあわせることで、様々な観点からの定量分析が可能となる。例えば「基本問題と応用問題」あるいは「出題予告した問題とそうでない問題」で正答率を比較するようなクロス集計分析が可能となる。また、各学生についてループリック観点別（学習到達目標別）に正答率を評価するようなことも可能となる。

以上のように、得点に縮約する以前の答案情報を用いた分析は、教科教育上の新たな知見や洞察を得

る一助になると考えられる。さらに、授業出席率や課題提出率、授業評価アンケート、関連科目成績のようなデータと組み合わせることで、更なる教育改善や学習支援が可能になると考えられる。

しかしながら、教員の手作業によって前処理、特に解答属性データを作成することは時間的にも労力的にも非常に大きな負荷となる。例として、採点済みの答案を使って、すべての解答の正答/誤答をデータ化する作業を考える。具体的には、表計算ソフト上に学生名簿を準備しておき、答案束を捲りながら正答であれば（赤マルがついていれば）1を、そうでなければ0を該当セルに入力していくことを考える。この作業は40名分の答案で1問あたり80~120秒の時間を要する（筆者が実際に作業して計測した時間に基づく、なお、時間のばらつきはミスによる手戻りに起因する）。1回分の見直し含めると、どんなに速くても2分は必要で、答案全体に対しては設問数を掛けて時間を見積もることができる。しかし、単調かつ単純な作業であるため、集中力が続かず、休憩が必要であったり、ミスによるやり直しがあったりするため、実際には想定以上の時間がかかり、精神的にもつらい作業となる。また、正誤属性のデータ化は採点マークの転記作業といえるが、それ以外の属性のデータ化は、その都度、解答記述を読み取っての判断が必要であり、さらに負荷は大きくなる。

一方で、表2のような設問属性データの作成は、答案束を捲るような作業は不要で、比較的短時間で終わることができる。

4. 前処理支援システムの設計と実装

手作業では多大な労力を必要とする前処理について、コンピュータを利用することで省力化・効率化を図ることができないかを検討した。

完全自動化という観点で前処理を考える場合、まずは、答案紙面から各設問の解答欄を自動識別させる必要がある。しかし、筆記解答方式の答案は、教員や教科によって書式が大きく異なるため、実務に耐える精度で同処理をすることは技術的に難しい。また、学生の解答記述に対して適切な属性を選定して付与することも技術的に困難がある。マルやバツの採点マークに基づく正誤属性や、記号解答属性（例えば「ア」「イ」「ウ」など）に限れば、人工知能による画像分類のアプローチも可能であるが、一般的な属性については教員による判断が不可欠となる⁹⁾。

以上のことから、我々は、完全自動化ではなく、

教員の作業を支援・補助するという観点で省力化と効率化を目指すことを考えた。具体的には、定型の繰り返し作業を自動化するツールや、作業に最適化して情報を整理して表示するツールなどを開発し、その利用を前提とした作業フローを提案した。以下に、その前処理の作業フローを示す。

- (1) 採点済みの答案束をスキャンして画像データ化する。
- (2) それを設問単位（解答欄単位）の画像にトリミング・分解して出力する。出力のイメージを図 1 に示す。
- (3) 分解出力された設問単位の解答記述画像を一覧表示し、解答属性を付与する。また、その情報を CSV ファイルに出力する。
- (4) 設問単位の CSV ファイルを集約し、答案全体の解答属性データとして整形して CSV ファイルに出力する。また、設問属性データの雛形を CSV ファイルに出力する。



図1 設問単位にトリミングした答案画像の出カイメージ

このフローに基づく作業のために、独立して動作する 3 つのツール「一括リネームツール」「一括トリミングツール」「属性情報付与ツール」と 1 つの Python スクリプト「データ集約スクリプト」を開発、リリースした。各ツールには、前処理に特化した GUI (グラフィカルユーザーインターフェース) と機能を実装した。また、環境構築やインストールをすることなく利用できる Windows アプリケーションとして開発した。以下、作業の流れに沿って、各ツールとスクリプトの詳細を説明する。

4.1 一括リネームツール

スキャンした答案画像に対して前処理をするためには、なんらかの方法で学生情報と画像を紐づける必要がある。ここでは、データ管理上の利便性を考え、ファイル名として画像に学生情報を紐づけるものとした。具体的には、画像ファイルに対して「01-高専太郎.jpg」「02-寝屋川次郎.jpg」「03-枚方三郎.jpg」のようなファイル名をつけて情報を管理する方式をとった。そして、この処理の補助のために図 2 に示すような一括リネームツールを開発した。ファイル

リネームのためのフリーソフトは多々存在するが、ここでは前処理に特化して次のような機能を実装した。1 点目は、改行区切りのテキストを使って変更後のファイル名を一括設定できる機能となる。これにより、エクセルなどにつくられている名簿からコピー&ペーストで新しいファイル名を一括して設定することができる。2 点目は、ツール画面上に画像の任意位置を拡大プレビュー表示できる機能となる。これにより、答案の氏名欄と変更後ファイル名の対応を簡単に確認することができる。

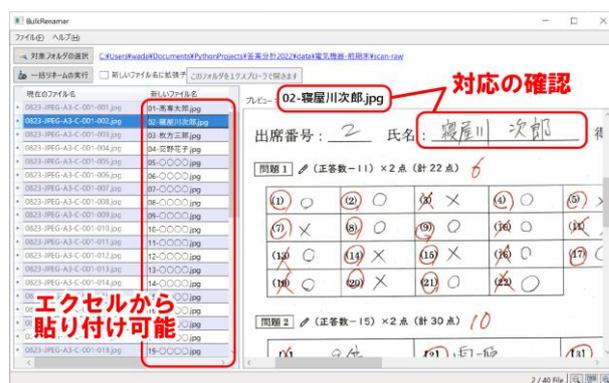


図2 一括リネームツールの実行画面

4.2 一括トリミングツール

本ツールは、答案全体画像を設問単位にトリミング (切り抜き) したものを一括して生成するツールとなる。実行画面を図 3 に示す。

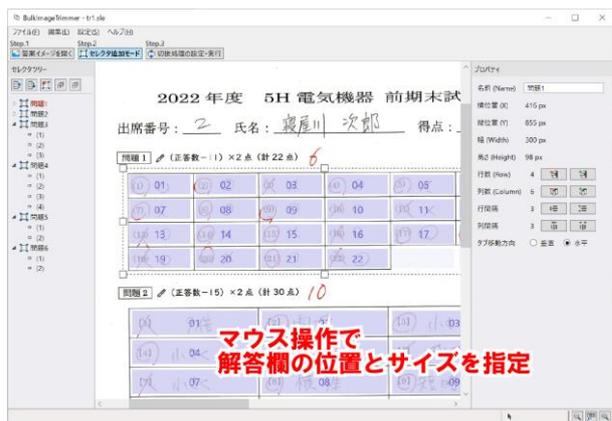


図3 一括トリミングツールの実行画面

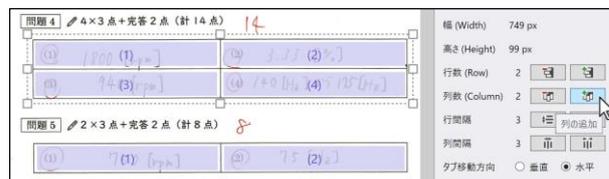


図4 解答欄領域を表形式で指定する例

ここでは、まず、トリミング位置と大きさを指示するための参照情報として、1枚の答案全体画像を読み込む。次に、マウスのドラッグ&ドロップ操作によって各解答欄の位置と大きさを指示していく。格子状に配置された解答欄に対しては、行数、列数、間隔（マージン）を指定して図4のように領域を指示することもできる。なお、各解答欄の領域には「問題1」「問題2」「問題5(1)」のように名前も設定する。これらが完了した後、一括処理を実行すると、指定フォルダ内にある全ての答案画像を対象にトリミング処理が一括適用される。そして、解答欄単位に分解された画像は、先に示した図1のように問題名つきのフォルダのなかに出力される。

4.3 属性情報付与ツール

本ツールは、解答欄単位に分解された画像を対象に、解答属性を付与するためのツールとなる。このツールでは属性を「タグ」、個々の解答にタグを付与する操作を「タグ付け」と呼称している。

解答に対するタグ付けは、設問ごとに行なう必要がある。そのために、まずは先のツールで生成されたフォルダから1つを指定して解答画像を読み込む。画像が読み込まれた状態の画面を図5に示す。



図5 属性情報付与ツールに解答画像を読み込んだ画面



図6 タグ(属性情報)の名前と外観の設定

ツールでは「正答」や「誤答」のような基本的なタグをプリセットとして用意しているが、分析の目的に応じて図6のようにタグの追加・削除・名前変更・外観変更ができるようになっている。解答に対するタグ付けは、基本的にマウスにより行なう。タグ付けが完了した状態を図7に示す。

タグ付けが完了した後は、その情報を図8に示すような解答属性データ(CSVファイル)として出力することができる。また、タグ付け作業の状態はJSONファイルとして保存でき、それを読み込めば作業を途中再開することもできる。



図7 解答にタグ付けがされた状態

	A	B	C	D	E	F	G	H	I
1 ID		正答	誤答	部分減点	空欄	判読困難			
2 01-高専太郎		1	0	0	0	0			
3 02-徳屋川次郎		0	1	0	0	0			
4 03-枚方三郎		0	1	0	0	0			
5 04-交野花子		1	0	0	0	0			

図8 属性情報付与ツールから出力された解答属性データ

4.4 データ集約スクリプト

先のツールからは、個々の設問に対する解答属性データが単体で出力される。しかし、分析フェーズでは、それらが表1のように1つに集約されていたほうが使いやすいことがある。また、特定の属性だけを抽出して結合したデータが望まれる場合もある(例えば、すべての設問に対する正答の属性だけを集約したデータなど)。

このようなデータの結合や抽出は、エクセルのような表計算ソフトでも可能である。しかし、一定の自動化も図れるため、その処理を行なうPythonスクリプトを作成した。先のツールのようにWindowsアプリケーションという形式にせず、スクリプト形式としているのは、柔軟性と拡張性を考慮したためである。例えば、どのような属性を抽出するかは分析ごとに変わるため、それに依って利用者が処理を柔軟に書き換えられるスクリプト形式をとっている。

なお、Python 言語を採用した理由は、データ分析用のライブラリが充実しており、分析フェーズにおいても Python が利用される可能性が高く、それとの親和性を考えたことにある。また、可読性に優れ、環境構築が比較的容易であることも理由に含まれる。なお、このスクリプトの実行には、Python 3.7 の実行環境と基礎知識が必要となる。

以上の各ツールとスクリプトは下記の URL において配布・公開している。

https://drive.google.com/drive/folders/1_55Bgn7uIj7TzW1lKZvZ_QOL7M3sLn-b

5. 実答案を使用した支援システムの評価

筆者の担当科目の定期試験答案を対象に、提案するフローに従って支援システムを使用した前処理（解答属性データの生成）に取り組んだ。各作業に要した時間を表 3 に示す。なお、リネーム作業には ADF（自動原稿送り装置）を使用した答案スキャンの時間（A3 片面カラー原稿 300dpi で約 1 秒/枚）を含む。また、タグは「正答」「誤答」「部分減点」「空欄」の 4 つを設定して付与するものとした。

表 3 提案法による前処理において各作業に要した時間

	試験1	試験2	試験3	試験4	試験5	試験6
受験者数	39	23	39	36	24	39
設問数	69	46	48	63	44	29
前処理対象	採点前	採点前	採点後	採点後	採点後	採点後
リネーム作業	8	4	4	4	3	4
トリミング作業	16	12	11	8	8	7
タグ付け作業	74	40	38	36	31	26
合計	98	56	53	48	42	37

* 各作業と合計の行は、作業に要した時間（単位：分）を表す

まず、2022 年度の前期中間で実施した試験 1~3 を対象に前処理を実施した（比較のために試験 1 と 2 では採点前の答案を使用した）。表 3 に示すように作業時間の大部分はタグ付け作業によるものとなった。特に、採点前答案では、教員が解答記述を見て正誤を判定しながらタグ付けするため、時間的にも労力的にも大きな負担となることが確認できた。なお、この評価実験のなかで、ツールの操作と表示についての新たなアイデアが得られたため、実験後、ツールに改良を加えている。

次に、同年度の前期末に実施した試験 4~6 を対象に同評価を実施した。試験 1~3 と比較して作業時間全般が短くなっているが、これは教員の作業慣れとツール改良による影響と考えられる。

ところで、タグ付けの作業時間は、基本的には受験者数と設問数の両方に比例して大きくなるが、実際には正答率などにも影響を受ける。例えば、全員が正答であれば、その設問に対するタグ付けは一瞬で完了する。このようなことから、試験 4 と 5 では受験者数×設問数に大きな違いがあるが、作業時間としては大差がついていない。

第 3 節では、手作業による前処理の時間を 1 問あたり約 2 分（採点済み答案 40 枚・正誤のみの属性取得）と見積もった。これと表 3 の結果を比較すると、提案法では（正誤以外の属性も含めて）手作業の場合の 30~60% の時間で前処理ができておりと評価できる。また、試験 1~6 の前処理において、各ツールは処理に必要な機能を提供し、異常終了やメモリリークすることなく安定して動作した。これより、ツール群が実務レベルで利用できる作業支援システムとなっていることを実証的に確認できた。

6. おわりに

本研究では、筆記解答方式の答案分析の前処理に対する作業フローの提案と、その作業支援システムの開発を行なった。また、提案した作業フローに沿って、約 40 名の採点済み答案から定量分析に必要な構造化データを概ね 1 時間未満で作成できることを実証的に確認した。このことは、答案分析に対するハードルを確実に下げたといえる。しかし、教員の多忙化が止まらないなかで、業務としてマストではない答案分析に対する心理的なハードルはまだ相当に高いと考えられる。この解消には、実際の答案分析によって得られた知見や洞察、効果などを教員同士で共有していくことが重要と考える。

本研究は JSPS 科研費 JP18K02886 の助成を受けたものである。

参考文献

- [1] 山田政寛, ラーニング・アナリティクス研究の現状と今後の方向性, 日本教育工学会論文誌, Vol. 41, No. 3, pp. 189-197, 2018
- [2] 藤本徹, 荒優, 山内祐平, 大規模公開オンライン講座 (MOOC) におけるラーニング・アナリティクス研究の動向, 日本教育工学会論文誌, Vol. 41, No. 3, pp. 305-313, 2018
- [3] 和田健, 早川潔, 谷野圭亮, 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, 日本高専学会第 27 回年会講演会講演論文集, C4-2, pp.75-76, 2021