

マルチモーダルLLMを用いた任意の帳票・用紙からの情報抽出システムの検討

佐々木風太[†] 陳思楠[†] 佐伯幸郎^{††} 中村匡秀[†]

[†] 神戸大学 〒657-8501 神戸市灘区六甲台町 1-1

^{††} 高知工科大学 〒782-8502 高知県香美市土佐山田町宮ノ口 185

E-mail: [†]safu@es4.eedept.kobe-u.ac.jp, ^{††}chensinan@gold.kobe-u.ac.jp, ^{†††}saiki.sachio@kochi-tech.ac.jp,
^{††††}masa-n@cmds.kobe-u.ac.jp

あらまし 手書き帳票のデジタル化において、従来のOCR技術は手書き文字認識精度の低さや複雑なレイアウトへの対応困難性が課題であった。近年、マルチモーダル大規模言語モデル（LLM）の発展により高精度な手書き文字認識が可能となり、先行研究では救急出動報告書読み取りシステムが高い性能を達成した。しかし、このシステムは救急出動報告書に特化した設計であり、異なる項目構成を持つ他の書類には適用できない汎用性の課題があった。本研究では、ユーザーが定義した項目に基づいてプロンプトを動的に生成する機能と、AIが書類構造を解析して項目を自動特定する機能により、任意の帳票・用紙から情報を抽出可能な汎用システムを提案する。評価実験では、履歴書、始末書、アンケート用紙という3種類の異なる業務書類に対して項目特定精度（F1スコア）100.0%、情報抽出精度（F1スコア）88.9%を達成し、書類種別に依存しない汎用的な情報抽出を実現した。

キーワード 手書き文字認識, 文書デジタル化, 大規模言語モデル, 文書処理, DX, OCR

Study on Information Extraction System for Arbitrary Forms and Documents Using Multimodal LLM

Futa SASAKI[†], Sinan CHEN[†], Sachio SAIKI^{††}, and Masahide NAKAMURA[†]

[†] Kobe University Rokkodai-cho 1-1, Nada-ku, Kobe, Hyogo 657-8501 Japan

^{††} Kochi University of Technology 185 Miyanokuchi, Tosayamada, Kami City, Kochi 782-8502, JAPAN

E-mail: [†]safu@es4.eedept.kobe-u.ac.jp, ^{††}chensinan@gold.kobe-u.ac.jp, ^{†††}saiki.sachio@kochi-tech.ac.jp,
^{††††}masa-n@cmds.kobe-u.ac.jp

Abstract In the digitization of handwritten forms, conventional OCR technology has faced challenges in low recognition accuracy for handwritten characters and difficulty in handling complex layouts. In recent years, the development of multimodal large language models (LLMs) has enabled high-accuracy handwritten character recognition, and prior research achieved high performance with an emergency dispatch report reading system. However, this system was designed specifically for emergency dispatch reports and had limitations in versatility, as it could not be applied to other documents with different item configurations. In this study, we propose a general-purpose system capable of extracting information from arbitrary forms and documents through a dynamic prompt generation function based on user-defined items and an automatic item identification function that analyzes document structure using AI. In evaluation experiments, we achieved item identification accuracy (F1-score) of 100.0% and information extraction accuracy (F1-score) of 88.9% for three different types of business documents: resumes, incident reports, and questionnaires, demonstrating versatile information extraction independent of document type.

Key words Handwritten character recognition, Document digitization, Large language model, Document processing, DX, OCR

1. はじめに

現代社会ではデジタルトランスフォーメーション (DX) の推進が急務となっており、様々な組織で業務効率化とコスト削減が求められている。しかし、手書き帳票や申請書類のデジタル化は依然として重要な課題である。従来の光学文字認識 (OCR) 技術は、印刷文字と比較して手書き文字の認識精度が著しく低く、特に日本語のひらがなや漢字の認識において顕著な限界があった [1]。また、複雑なレイアウトや多様な入力形式を持つ書類では、項目の境界認識や情報抽出が困難である。

近年、マルチモーダル大規模言語モデル (LLM) の発展により、画像に含まれる文字情報の認識・抽出に革新的な可能性が示されている [2]。これらのモデルは、文脈理解能力と高度な画像処理能力を兼ね備え、複雑なレイアウトに対しても高い適応性を示す。先行研究で開発された救急出動報告書読み取りシステム [3],[4] は、Google Gemini 2.5 Pro を用いて平均文字誤り率 4.76% を達成し、高いレベルでの手書き文書処理能力を実現した。

しかし、このシステムは救急出動報告書に特化した設計であり、プロンプト内で「出動番号」「傷病者氏名」「バイタルサイン」など救急特有の項目が固定的に定義されている。このため、履歴書、見積書、申請書といった異なる項目構成を持つ他の書類には適用できない汎用性の課題があった。実際の組織では、顧客申込書、診断書、各種申請書、契約書、アンケート用紙など多様な手書き書類を日常的に扱っており、書類種別ごとに専用システムを開発・運用することはコストと労力の面で非効率である。

本研究の目的は、任意の帳票・用紙から情報を抽出可能な汎用システムを構築することである。キーアイデアは、ユーザーが定義した項目に基づいてプロンプトを動的に生成し、マルチモーダル LLM に情報抽出を実行させることである。これを実現するため、(1) ユーザーが抽出したい項目を自由に定義し、最適化されたプロンプトを自動生成する動的プロンプト生成機能、(2) AI が書類構造を解析して抽出すべき項目を自動特定する項目特定機能の 2 つを提案する。

評価実験では、履歴書、始末書、アンケート用紙という 3 種類の異なる業務書類に対して、項目特定精度 (F1 スコア) 100.0%、情報抽出精度 (F1 スコア) 88.9% を達成した。特にテキスト入力形式の書類 (履歴書、始末書) では F1 スコア 100.0% を達成し、書類種別に依存しない汎用的な情報抽出を実現した。

2. 準備

2.1 マルチモーダル LLM

近年、テキスト、画像、音声、動画などの複数のモダリティを統合的に処理するマルチモーダル大規模言語モデルが急速に発展している。OpenAI の GPT-4V、Anthropic の Claude、Google の Gemini など、様々なモデルが開発されており、従来の単一モダリティモデルとは異なり、文字、画像、音声などの異なる種類の情報を同時に処理し、それらの関係性を理解することができる。

これらのモデルは、ユーザーが指定したプロンプト (指示文) に基づいて、画像に含まれる情報を構造化して抽出することが可能である。特に手書き文字認識においては、従来の OCR 技術と比較して文脈理解能力と高度な画像処理能力を有し、複雑なレイアウトや多様な筆記スタイルに対してより柔軟な対応が期待される。

先行研究では、手書き医療文書の認識精度を重視して、現在利用可能な主要なマルチモーダルモデルの性能評価を実施した。評価対象として、OpenAI o3 [5]、Claude Opus 4 [6]、および Google Gemini 2.5 Pro [7] を選定し、図 1 に示す救急出動報告書の文章部分を切り取った画像 3 枚を用いた認識精度の比較を行った。

評価指標として文字誤り率 (Character Error Rate: CER) を採用し、認識結果と正解テキストの文字レベルでの比較により精度を算出した。実験結果を表 1 に示す。

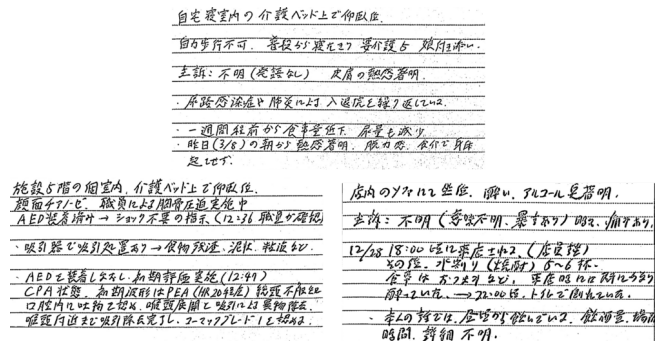


図 1: 救急出動報告書の切り取り画像 (上部: 画像 1, 下部左: 画像 2, 下部右: 画像 3)

表 1: マルチモーダル大規模言語モデルの性能比較

モデル	画像 1	画像 2	画像 3	平均
OpenAI o3	24.24	23.12	31.25	26.20
Claude Opus 4	47.73	21.61	46.25	38.53
Gemini 2.5 Pro	3.03	2.51	8.75	4.76

※数値は文字誤り率 (CER) を % で表示

表の結果から、OpenAI o3 では平均 26.20%、Claude Opus 4 では平均 38.53% の文字誤り率となったのに対し、最も優れた性能を示したのは Gemini 2.5 Pro であり、平均 4.76% という極めて低い文字誤り率を達成した。特に画像 2 では 2.51% という高精度を実現している。

この結果から、Gemini 2.5 Pro は他のモデルと比較して顕著に高い認識精度を示した。

2.2 救急出動報告書読み取りシステム

前節で述べた Gemini 2.5 Pro の高い認識性能を活用し、先行研究では救急医療現場における手書き文書デジタル化を目的とした救急出動報告書読み取りシステムが開発された [3],[4]。

2.2.1 システムの概要

図 2 に示す救急出動報告書は、救急隊員が現場で取得した患者の症状、応急処置、周辺環境など、多種多様な情報を包括的に記録する文書である。出動番号、傷病者氏名、バイタルサイン

ン(血圧, 脈拍, 体温など), 接触時の状況, 応急処置の内容, 搬送先医療機関など, 後続の医療対応や症例検討に不可欠な情報が記載されている。しかし, 現状ではこうした情報は紙ベースで記録されており, 帰署後に隊員や事務担当者が手作業でシステムへ入力する必要がある。この作業は1件あたり約15分を要し, 出動件数の増加に伴い組織全体の事務負担を増大させていた。

このシステムは救急出動報告書の項目構成とレイアウトに特化して設計されており, プロンプト内で「出動番号」「傷病者氏名」「バイタルサイン」など救急出動報告書固有の項目が明示的に定義されている。このため, 履歴書, 見積書, 申請書など, 異なる項目構成を持つ他の種類の手書き書類への適用ができない。実際の組織では複数種類の手書き書類を日常的に扱っており, 書類種別に依存しない汎用的な情報抽出システムが求められている。

3. 提案手法

3.1 目的とアプローチ

本研究の目的は, 任意の帳票・用紙から情報を抽出可能な汎用システムを構築することである。

本研究のキーアイデアは, ユーザーが定義した項目に基づいてプロンプトを動的に生成し, マルチモーダル LLM に情報抽出を実行させることである。これにより, 書類種別に依存しない汎用的な情報抽出を実現する。

この目的を達成するため, 満たすべき要件を以下に示す。

RQ1: 任意の書類種別に対応可能な動的プロンプト生成

ユーザーが抽出したい項目を自由に定義でき, その項目定義に基づいて最適化されたプロンプトを自動生成する機能を実現する。これにより, 書類種別ごとにシステムを開発し直すことなく, 多様な帳票に対応できるようにする。

RQ2: 項目定義作業を支援する AI 自動特定機能

RQ1 を実現する上で, ユーザーが項目を手動で定義する作業は負担となる。そこで, AI が書類画像を解析して抽出すべき項目を自動的に特定する補助機能を実現する。これにより, 項目定義の手間を削減し, RQ1 の実用性を高める。

上記の各要件に対してのアプローチを以下に示す。

3.2 A1: 動的プロンプト生成機能

この機能は, RQ1 を満たすための中核機能であり, ユーザーが定義した項目情報からマルチモーダル LLM 用のプロンプトを自動生成する。従来システムでは項目が固定的に記述されていたが, 本機能では任意の項目定義からプロンプトを動的に構築することで, 書類種別に依存しない汎用的な情報抽出を実現する。

ユーザーは, Web インターフェース(図3)を通じて, 項目名, データ型, 必須/任意, 説明などを入力する。システムは定義された項目情報に基づいて, 基本指示文の構築, 項目別指示の追加, JSON 形式での出力指定, 例外処理の定義を行い, プロンプトを自動生成する(図4)。

生成されたプロンプト設定は, プリセット(プリセット名, 書類種別の説明, 項目定義リスト, プロンプトテンプレート)として保存され, 同種の書類を処理する際に再利用できる。

3.3 A2: 項目特定機能

この機能は, RQ2 を満たすための補助的機能であり, A1 における項目定義作業を支援する。ユーザーが書類画像をアップロードすると, マルチモーダル LLM (Gemini 2.5 Pro) が書類を解析し, レイアウト認識, 項目識別, データ型推定を実行して項目定義を自動生成する。

図 2: 救急出動報告書

このシステムは, ユーザーが報告書画像をアップロードするだけで, Gemini 2.5 Pro による高精度な手書き文字認識を実行し, 報告書に記載された情報を構造化データとして自動抽出する。システムの特徴は, 救急出動報告書に特化したプロンプト設計により, 報告書内の項目を自動的に抽出・構造化できる点にある。プロンプト内で「出動番号」「傷病者氏名」「バイタルサイン」などの報告書固有の項目が明示的に定義されており, これにより手作業での入力を大幅に削減できる。

評価実験では, 救急出動報告書の 26 項目について平均 83.7% の完全一致率を達成し, 従来の手作業による転記時間 15 分を大幅に短縮することで, 高いレベルでの手書き文書処理能力を実証した [4]。

2.2.2 着目する課題

しかしながら, 先行研究のシステムには以下の課題が存在する。

課題 P: 書類種別に特化したシステム設計



図 3: 項目定義のユーザーインターフェース

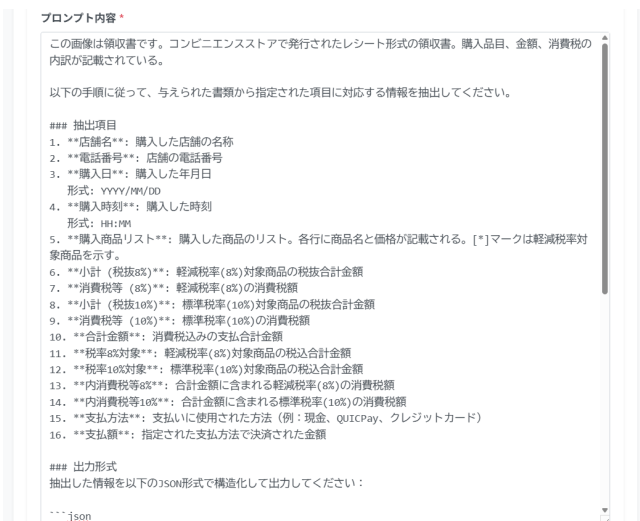


図 4: 動的に生成されたプロンプトの例

図 5 に示すように、AI が自動特定した項目リストが表示され、ユーザーは項目の追加・削除・編集が可能である。確定後は A1 の動的プロンプト生成機能に渡され、プロンプトの自動生成とプリセット保存が実行される。この機能により、初めて扱う書類でも効率的にプリセットを作成でき、RQ1 の実用性を高めることができる。

3.4 A3: プリセット管理と書類解析機能

この機能は、RQ1 を支援する機能であり、A1 で生成された項目設定を再利用可能な形で保存・管理する。保存されたプリセットは編集・更新・削除が可能であり、組織内で共有することで項目定義を統一できる。

保存されたプリセットを使用した書類解析では、(1) 書類画像のアップロード、(2) プリセット選択、(3) 項目定義の読み



図 5: 項目特定機能のユーザーインターフェース

込み、(4) A1 によるプロンプト自動生成、(5) LLM による情報抽出、(6) JSON 形式での結果出力、という処理フローで実行される。

この機能により、組織で定期的に扱う定型書類に対して、初回のみ項目定義（手動入力または A2 による自動特定）を行えば、2 回目以降はプリセット選択だけで自動的に情報抽出が完了する。従来の手作業と比較して、大幅な時間短縮が実現される。

4. 評価実験

提案システムの有効性を検証するため、異なる種類の手書き書類を用いた情報抽出精度の評価実験を実施した。

4.1 実験設定

実験では、図 6～図 8 に示す 3 種類の手書き書類を使用した。これらはそれぞれ異なるレイアウト、項目構成、記入形式を持つ業務書類のサンプルであり、提案システムの汎用性を検証するのに適している。

評価対象とした書類の特徴は以下の通りである：

- 履歴書 (図 6)：氏名、生年月日、住所、学歴、職歴、資格、自己 PR など多様な項目を含む一般的な求職用書類
 - 始末書 (図 7)：宛先、氏名、日付、標題、概要、原因、対策などの自由記述を含む社内文書
 - アンケート用紙 (図 8)：性別・年代の選択、来店のみつけ、利用頻度、満足度評価 (5 段階評価で該当項目に○をつける形式)、自由記述など多様な回答形式を含む顧客アンケート
- 各書類について、手動で作成した正解データを基準として評価を行った。

4.2 評価方法

提案システムの有効性を検証するため、以下の 2 つの観点から評価を実施した。

4.2.1 項目特定機能の評価

AI による自動項目特定機能の性能を評価した。評価手順は以下の通りである：

- 各書類について、人手で正解となる項目リストを作成
- システムに書類画像を入力し、AI に項目を自動特定

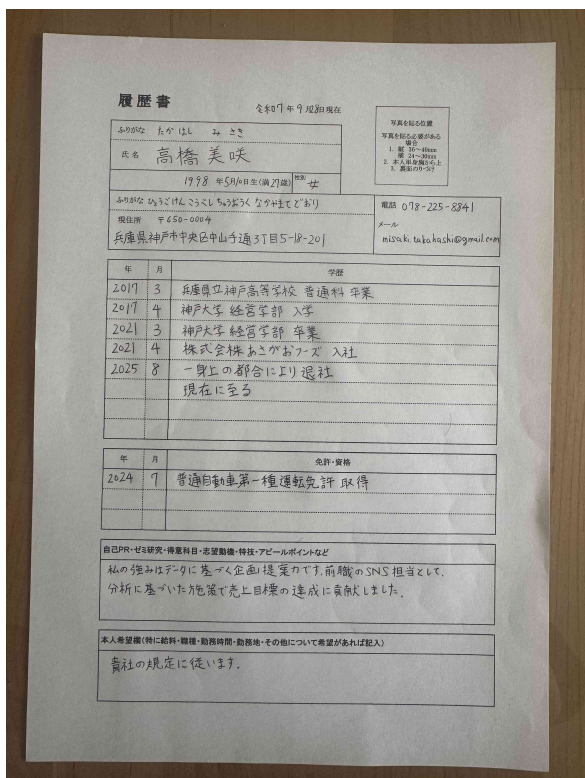


図 6: 履歴書 (評価実験書類 1)

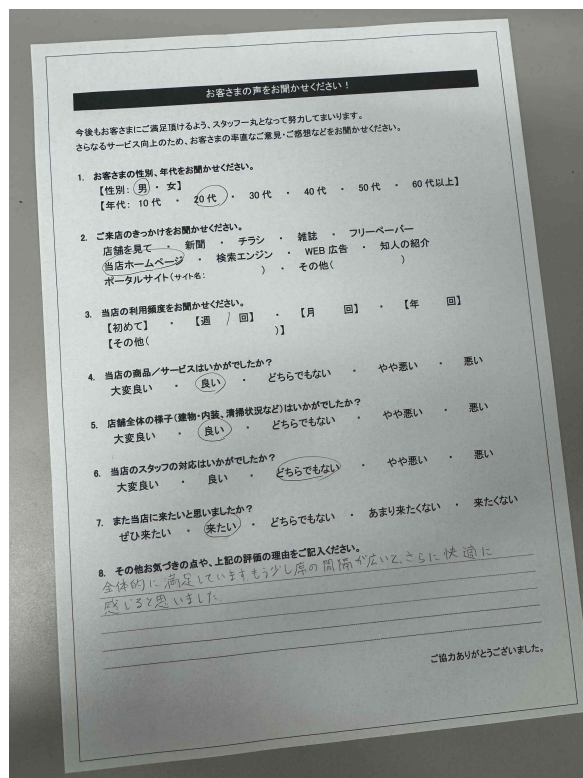


図 8: アンケート用紙 (評価実験書類 3)

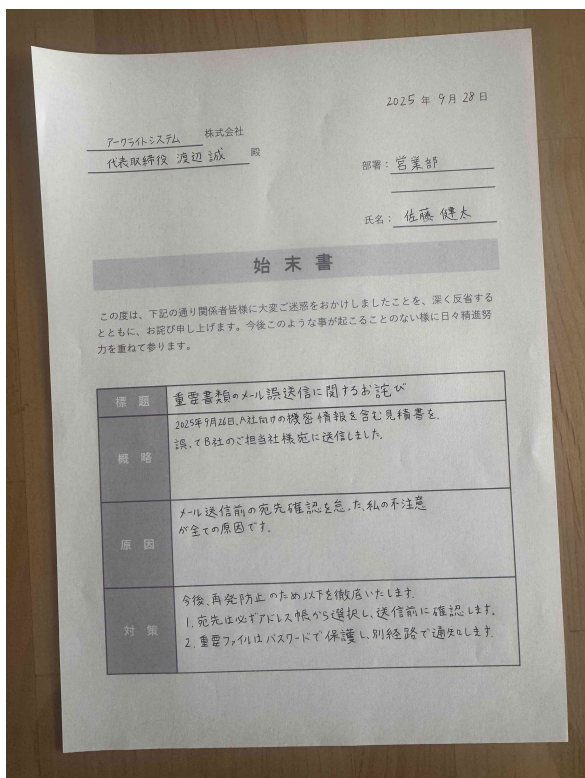


図 7: 始末書 (評価実験書類 2)

させる

(3) システムが特定した項目リストと正解の項目リストを比較し、適合率・再現率・F1 スコアを算出

評価指標として、情報検索分野で広く用いられる以下の3つの指標を採用した:

- **適合率 (Precision)**: 抽出された情報のうち、正解と一致した情報の割合
 - **再現率 (Recall)**: 正解情報のうち、システムが正しく抽出できた情報の割合
 - **F1 スコア**: 適合率と再現率の調和平均。両者のバランスを表す総合指標
- 表 2 に項目特定精度の評価結果を示す。

表 2: 項目特定機能の評価結果

書類種別	適合率 (%)	再現率 (%)	F1 スコア (%)
履歴書	100.0	100.0	100.0
始末書	100.0	100.0	100.0
アンケート用紙	100.0	100.0	100.0
平均	100.0	100.0	100.0

全ての書類種別において、AI は書類に存在する全ての項目を正確に特定し、適合率・再現率・F1 スコアともに 100.0% を達成した。

4.2.2 情報抽出精度の評価

次に、動的プロンプト生成機能による情報抽出精度を評価した。評価手順は以下の通りである:

- (1) 各書類について、人手で各項目の正解値を作成
- (2) 項目特定機能で特定された項目に基づき、動的プロンプト生成機能でプロンプトを自動生成
- (3) 生成されたプロンプトを用いて、マルチモーダル LLM に書類画像から情報を抽出させる
- (4) 抽出された各項目の値と正解値が完全に一致したもののみを正解として、適合率・再現率・F1 スコアを算出

表 3 に情報抽出精度の評価結果を示す。

表 3: 情報抽出精度の評価結果

書類種別	適合率 (%)	再現率 (%)	F1 スコア (%)
履歴書	100.0	100.0	100.0
始末書	100.0	100.0	100.0
アンケート用紙	66.7	66.7	66.7
平均	88.9	88.9	88.9

履歴書、始末書については、適合率・再現率・F1 スコアともに 100.0% を達成し、全ての項目を完全に抽出できた。一方、アンケート用紙では F1 スコア 66.7% にとどまった。抽出に失敗した 3 項目は、設問 1 (性別・年代の選択, ○印形式), 設問 3 (利用頻度の記述), 設問 6 (スタッフの対応評価, ○印形式) であった。全体平均では F1 スコア 88.9% (適合率 88.9%, 再現率 88.9%) を達成した。

4.3 考察

評価実験の結果、提案システムは項目特定機能において 100.0% の精度を達成し、多様な書類種別に対して抽出すべき項目を完全に自動特定できることが確認された。情報抽出精度については、テキスト入力形式の書類 (履歴書、始末書) で F1 スコア 100.0% を達成したが、選択式を含むアンケート用紙では 66.7% にとどまり、平均 F1 スコア 88.9% となった。

アンケート用紙での精度低下の原因を分析した結果、抽出に失敗した 3 項目のうち 2 項目 (設問 1 の性別・年代, 設問 6 のスタッフ対応評価) は選択式 (○印をつける形式) であり、手書き○印の認識困難性が主な要因であった。また、設問 3 (利用頻度) は記述式項目であるが、「週/回」「月 回」「年 回」という複数の単位を含む複雑な記入形式であり、手書き数字と単位の組み合わせの認識が不完全であった。

一方、テキスト入力形式の書類 (履歴書、始末書) では、従来システムのような書類固有の設定なしに完全な精度での抽出を実現し、提案システムの高い汎用性が確認された。

今後の課題として、画像前処理による○印領域の強調や、選択肢の位置関係を考慮したプロンプト設計、複雑な記入形式を持つ項目への対応強化などの改善が必要である。

5. まとめ

本研究では、従来の手書き帳票処理システムが抱える書類種別特化の制約を克服し、任意の帳票・用紙から情報を抽出可能な汎用システムを提案した。

提案システムでは、動的プロンプト生成機能により、ユーザーが定義した項目に基づいて最適化されたプロンプトを自動生成することで書類種別に依存しない柔軟な情報抽出を実現した。また、AI による自動項目特定機能により、マルチモーダル LLM を活用した書類構造の自動解析によって抽出すべき項目を完全に特定し、ユーザーの手動設定負担を大幅に軽減した。

評価実験では、履歴書、始末書、アンケート用紙という 3 種類の異なる業務書類に対して項目特定精度 100.0%、情報抽出精度 (F1 スコア) 88.9% を達成した。テキスト入力形式の書

類 (履歴書、始末書) では F1 スコア 100.0% を達成し、書類種別に依存しない汎用的な情報抽出が可能であることが示された。一方、アンケート用紙では F1 スコア 66.7% にとどまり、抽出に失敗した 3 項目のうち 2 項目は選択式 (○印形式) であり手書き○印の認識困難性が、1 項目は複雑な記入形式を持つ記述式項目であり単位を含む数字の認識が課題として明らかになった。

今後は、選択式質問における認識精度の向上、対応書類種別の拡大、実運用環境での検証を通じて、より広範囲での実用化を目指す。

謝辞 本研究の一部は JSPS 科研費 JP25H01167, JP25K02946, JP25K24389, JP24K02765, JP24K02774, JP23K17006, JP23K28091, JP23K28383 の研究助成を受けて行われている。本研究は神戸市消防局との共同研究の一環で行われている。

文 献

- [1] “tesseract を試しに使用して手書き文字 (日本語) の認識精度を確認してみた,” <https://qiita.com/smiler5617/items/b7efafd5f3f6b0261296>, accessed October 21, 2025.
- [2] “生成 AI 搭載の OCR とは? 従来の OCR との 3 つの違いや事例 4 選も紹介,” https://metaversesouken.com/ai/generative_ai/ocr/#4000, accessed October 21, 2025.
- [3] 佐々木風太, 陳思楠, 中村匡秀, “生成 AI を活用した救急出動報告書読み取りシステムの研究開発,” 電子情報通信学会技術研究報告, vol.124, no.334, LOIS2024-66, pp.84–89, January 2025.
- [4] 佐々木風太, 佐伯幸郎, 中村匡秀, “生成 AI を用いた手書きフォームの高精度なデータ認識・抽出手法の検討,” 電子情報通信学会技術研究報告, vol.125, no.115, SS2025-20, pp.116–121, July 2025.
- [5] “OpenAI o3 and o3-mini,” <https://openai.com/index/openai-o3-and-o3-mini/>, accessed October 21, 2025.
- [6] “Introducing Claude Opus 4,” <https://www.anthropic.com/news/claude-opus-4>, accessed October 21, 2025.
- [7] “Gemini 2.5: Our most intelligent AI model,” <https://blog.google/technology/google-deepmind/gemini-model-thinking-updates-march-2025/>, accessed October 21, 2025.