P2-05

ワードプロセッサのスキル評価のための レポートとそのチェックリストを用いた生成 AI による自動評価の検討

Investigating Automatic Evaluation with Generative AI Using Reports and Checklists for Skill Assessment of Word-processor

久保田 真一郎[*1],
Shin-Ichiro KUBOTA[*1],
[*1]熊本大学
[*1]Kumamoto University
Email: kubota@cc.kumamoto-u.ac.jp

あらまし：ICT リテラシーを対象とした授業においてレポート課題を評価するためのチェックリストを準備し，生成 AI に対してレポート課題内容および正解の例，チェックリストを入力とする場合の自動評価の検討を行った．また，学習者向けの応用を想定してヒント生成を試みた．
キーワード：自動評価，チェックリスト，生成 AI

1. はじめに

大学初年次学生約 1800 名を対象に ICT リテラシーを学ぶ授業を展開している．当該授業は，1 週間に 1 回の授業で，15 回行われ，途中 4 つの課題が課せられる．提出されるレポートやスプレッドシートなどの課題ファイルを評価するコストは膨大である．

2020 年から生成 AI を利用した業務改善などの取り組みが様々行われている．OpenAI のサービス ChatGPT[(1)] は大規模言語モデルを利用して人間のような応答をテキスト文で生成するサービスで，Few-shot や Chain-of-Thoughs などの手法で問い合わせることで，学習文脈にあわせた学習プロセスなどについて応答を返す．これらを応用した教育実践の検討が存在する[(2)]．

本研究では，提出された課題ファイルを評価するコストを削減するために，生成 AI に課題ファイルを評価する観点を記述したチェックリストを入力して，自動的な評価ができないか検討する．

2. 研究方法

今回は，ICT リテラシーの授業で扱う 4 つの課題のうち，ワードプロセッサをテーマとした回の課題ファイルを対象として検討した．この授業は 9 名の教員で担当しているが 9 名で手分けして評価のばらつきが発生しないように，レポート課題を評価する観点をチェックリストとしてまとめている．レポート課題の内容およびレポートのチェックリスト，レポートの正解例を PDF ファイルで作成し，OpenAI の ChatGPT に API 経由で入力し，チェックリストの項目ごとに判定結果を出力する構成を試みた．

チェックリストの一部を表 1 に示す．ICT リテラシーを学ぶ授業のため，評価する観点はいずれもレポートのレイアウトに関する観点となっている．そのため，生成 AI に対してレイアウトの情報を入力する必要がある．そこで HTML によって文章構造を反映させて入力することでレイアウトの判定ができるか検討を行なった．

表 1 レポートを評価するチェックリストの一部

チェック項目
（省略）
レポートのタイトルを書いている。文字はゴシック体のように太めの文字で、文字サイ ズは本文の文字サイズより大きい 14 ポイント程度で、中央寄せで書いている。
所属、学生番号、名前を書いている。文字と文字サイズは本文と同じで、中央寄せなど でわかりやすく書いている。タイトルと所属との間は 1 行空いている。
（省略）

自動評価のための処理の流れを図 1 に示す．まず，入力に必要なファイルは，評価対象となる学習者が提出したレポートの PDF ファイル，そしてレポート課題の内容が記載された PDF ファイル，チェックリストの PDF ファイル，レポートの正解例が記述された PDF ファイルの 4 つのファイルである．これらのファイルを ChatGPT の API サービスに入力するライブラリ LangChain[(3)] の PDF ファイルを扱う機能を使った．LangChain により PDF ファイル内のテキストを扱うことができるが，その構造は HTML に変換する必要があり，いくつかの PDF から HTML に変換するライブラリを試したが，ChatGPT の API サービスを使い，HTML に変換するプロンプトとともに入力することで HTML にある程度構造化できた．しかし，その精度を検討するまでは至っておらず，精度の検討が必要である．レイ

アウトを評価するのであれば画像で扱う検討も必要と考えている．また，レポート内の画像ファイルの扱いについて検討が必要である．

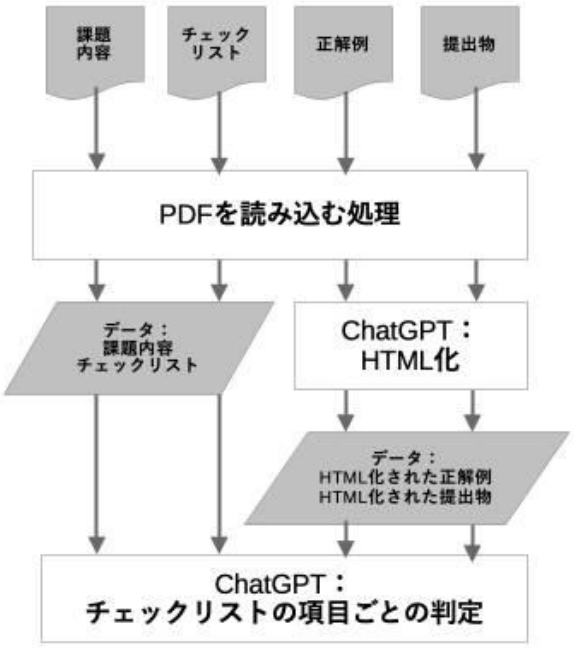

図 1 処理の流れ

課題内容のテキストデータおよびチェックリストのテキストデータ，正解例を HTML 化したデータ，提出されたレポートを HTML 化したデータを入力として，システムプロンプトとユーザプロンプトを構成し，ChatGPT の API へ入力した結果を得た．システムプロンプトは図 2 のように構成し，ユーザプロンプトは図 3 のように構成した．{{assignment}}に対応するようにレポート課題の内容，{{checklist}}に対応するようにチェックリストの内容，{{sample}}に対応するように HTML 化された正解例，{{submission}}に対応するように HTML 化された提出物をプロンプトの最後に追加している．

あなたは初年次の大学生に対してワードプロセッサの使い方を教える大学の教員です．{{assignment}}はあなたが課したレポートの内容です．{{checklist}}はレポートを評価するためのチェックリストです．評価は OK か NG か NA のいずれかです．あなたは特に文章と参考文献との対応関係を示す記号が文章中にあるかチェックします．チェックリストのうち用紙サイズや文字の大きさなどの見た目を評価できません．評価できない項目は NA と判定します．{{sample}}は正解の例です．{{sample}}はチェックリストのすべての項目を満たします．

図 2 システムプロンプト

{{submission}}は学習者が提出したレポートです．チェックリストを使ってチェックリストの項目ごとに OK か NG かレポートを判定してください．NG の場合のみ OK にするためのヒントを出力してください．

図 3 ユーザプロンプト

複数実行してみたが，うまく判定する項目と判定できない項目があり，その傾向を整理した上で，プロンプトの構成を検討する必要がある．また，プロンプトの構成方法を先行事例を参考に組み合わせ，結果を比較しながら分析的に構成する必要がある．また，ChatGPT のモデルは 3.5-Turbo と 4.0-Turbo で比較を行い，検討する必要がある．

3.　まとめ

ICT リテラシーを対象とした授業においてレポート課題を評価するためのチェックリストを準備し，生成 AI に対してレポート課題内容および正解の例，チェックリストを入力とする場合の自動評価の検討を行った．チェック項目がレイアウトを評価する内容になっているため，文章の構造を HTML 化することで評価したが，画像による評価，HTML 化せずそのままの評価など精度を比較しながら検討が必要である．また，システムプロンプトとユーザプロンプトのそれぞれの構成についても先行事例をもとに精度を比較しながら検討が必要である．ChatGPT のモデルによる違いも検討が必要と考えている．

参考文献

(1) OpenAI, https://openai.com/（2025 年 6 月 4 日 確認）
(2) Phung, T., Pădurean, V.A., Singh, A., Brooks, C., Cambronero, J., Gulwani, S., Singla, A. and Soares. G., “Automating Human Tutor-Style Programming Feedback: Leveraging GPT-4 Tutor Model for Hint Generation and GPT-3.5 Student Model for Hint Validation”, In Proceedings of the 14th Learning Analytics and Knowledge Conference (LAK '24)（2024）
(3) LangChain, https://www.langchain.com/（2025 年 6 月 4 日 確認）