

も く じ

■開催日時：2023年7月15日（土）

■テーマ：ICT を活用した学習支援と教育の質保証／一般

- 1) 生成型 AI を用いた学習アドバイジングの提案及び評価-----1
○釣部勇人(公立千歳科学技術大学大学院)，高野泰臣(公立千歳科学技術大学大学院)，
上野春毅(公立千歳科学技術大学)，小松川浩(公立千歳科学技術大学)
- 2) 生成系 AI との相談を取り入れた数理データサイエンス AI 教育プログラムの実践と検証-----6
○二瓶裕之(北海道医療大学)，西牧可織(北海道医療大学)，米田龍大(北海道医療大学)
- 3) 看護過程での臨床判断支援システムにおける大規模言語モデルを活用したフィードバックの自動生成-----14
○佐藤玲央(公立千歳科学技術大学)，木口幸子(北海道文教大学)，辻幸美(北海道文教大学)，
初見温子(北海道文教大学)，藤長すが子(北海道文教大学)，高岡哲子(北海道文教大学)，
山川広人(公立千歳科学技術大学)，小松川浩(公立千歳科学技術大学)
- 4) プログラミングロボットを用いたグループ作品制作におけるメンバーの役割分担の現状と課題-----19
○布施泉(北海道大学)，野口孝文(北海道大学)
- 5) 故障予測および作業内容予測機能を持つカスタマエンジニアの学習支援システムの提案-----24
○中村伊吹(公立諏訪東京理科大学)，左座祐之助(公立諏訪東京理科大学)，
増田武史(公立諏訪東京理科大学)，広瀬啓雄(公立諏訪東京理科大学)，三代沢正(公立諏訪東京理科大学)
- 6) Scratch ブロックの多様な組み合わせを考慮したプログラム自動採点方法の開発 -----32
○畠中明哉(公立諏訪東京理科大学大学院)，広瀬啓雄(公立諏訪東京理科大学大学院)，
尾崎剛(公立諏訪東京理科大学大学院)
- 7) 野菜摂取における講義とベジチェックがもたらす大学生の健康行動変容への影響
～スタディエンゲージメントと自己効力感の調査から～-----38
○柳谷泰志(大阪府立大学)，阿部祐来(大阪公立大学大学院)，真嶋由貴恵(大阪公立大学大学院)，
梶田聖子(大阪公立大学大学院)
- 8) AI が作成したソーシャルワーク実習記録の考察文章の評価
～学生が作成した文章との比較分析～-----43
○坂本毅啓(北九州市立大学)，佐藤貴之(北九州市立大学)，中原大介(福山平成大学)

9) 大学1年生の生成AI利活用に関するアンケート調査報告-----	51
○森祥寛(金沢大学)	
10) 学習支援ロボットのロールによるセルフレビューへの影響-----	57
○佐田竣祐(電気通信大学), 柏原昭博(電気通信大学)	
11) インタラクティブロボット講義におけるケーススタディ分析と講義インタラクショモデルデザイン----	65
○島崎俊介(電気通信大学), 柏原昭博(電気通信大学)	
12) 心拍変動の計測によるビジュアル型言語とテキスト型言語の学習時の精神的ストレス分析について-----	73
○梅澤克之(湘南工科大学), 越川拓海(旭情報サービス), 中澤真(会津大学短期大学部), 平澤茂一(早稲田大学)	
13) 線形代数授業におけるオンライン問題の活用と学習分析-----	81
○吉富賢太郎(大阪公立大学)	
14) ドローイングストロークの線形回帰に基づくパース技術評価に関する基礎的検討-----	87
○杉井奏斗(信州大学), 永井孝(ものづくり大学), 香山瑞恵(信州大学)	
15) 顔検出と教員の発話による授業中の学習者の集中度評価システムの開発-----	94
○左座祐之助(公立諏訪東京理科大学大学院), 中村伊吹(公立諏訪東京理科大学大学院), 尾崎剛(公立諏訪東京理科大学大学院), 広瀬啓雄(公立諏訪東京理科大学大学院)	
16) 呼吸活動での胸郭運動計測の試み-----	99
○伊藤敏(岐阜聖徳学園大学), 井上祥史(岩手大学), 鷲野嘉映(愛知みずほ短期大学)	
17) 情報活用学士力を測定するメトリクス導出のための一検討-----	104
○小俣昌樹(山梨大学), 石田雪也(公立千歳科学技術大学), 金子大輔(北星学園大学), 吉川雅修(山梨大学), 古賀崇朗(佐賀大学)	
18) 人間的力育成科目の体系化に向けた共通性と個別性の抽出-----	112
○村松慶一(早稲田大学), 松居辰則(早稲田大学)	
19) L-VEIGe: 誤答画像生成を活用した語彙学習支援システムー画像の示唆性に対する指標の検討-----	118
○杉田一樹(北陸先端科学技術大学院大学), 太田光一(北陸先端科学技術大学院大学), 谷文(北陸先端科学技術大学院大学), 長谷川忍(北陸先端科学技術大学院大学))	

20) 英語教育支援システム Zoroe を用いたディクテーション分析に基づくリスニング指導-----	125
○照井雅子(近畿大学), 溝渕昭二(近畿大学)	
21) 初修外国語向け HyFlex 型授業デザインの最適化ー日本の三大学における実践報告ー-----	129
○杉江聡子(札幌国際大学), 王松(関西外国語大学短期大学部), 曲明(室蘭工業大学), 砂岡和子(早稲田大学)	
22) タブレット端末を用いた認知症スクリーニング検査における自動採点機能の実装-----	137
○倉持宏斗(東京工業高等専門学校), 小久保奈緒美(量子科学技術研究開発機構), 吉本定伸(東京工業高等専門学校)	
23) 没入型 VR を利用した大量調理シミュレータの開発	
ー自己観察機能 (ミラーリング機能) を利用した練習システムの検討-----	140
○中山 洋(東京電機大学) 堀端 薫(女子栄養大学)	
24) 学習履歴と教員ポリシー双方を考慮した CBT の難易度分類アルゴリズムの構築-----	142
○荒関虹希(公立千歳科学技術大学大学院), 上野春毅(公立千歳科学技術大学), 小松川浩(公立千歳科学技術大学)	

生成型 AI を用いた学習アドバイジングの提案及び評価

釣部 勇人^{*1}, 高野 泰臣^{*1}, 上野 春毅^{*2}, 小松川 浩^{*1}

^{*1} 公立千歳科学技術大学大学院 理工学研究科, ^{*2} 公立千歳科学技術大学 理工学部

A Proposal and Evaluation of Learning Advising using a Generative AI

Taketo Tsurube^{*1}, Yasuomi Takano^{*1}, Haruki Ueno^{*2}, Hiroshi Komatsugawa^{*1}

^{*1} Graduate School of Science and Engineering, Chitose Institute of Science and Technology

^{*2} Faculty of Science and Technology, Chitose Institute of Science and Technology

In this paper, we propose a method for automatically generating learning advice using a large language model, and evaluate that. In our previous study, advising message were created for each cluster based on learner classification using online learning history data. In the present study, we indicate that introducing a generative model allows us to create automatic advice using a generative AI that is combined with the learner characteristics within the cluster.

キーワード: 自然言語処理, 人工知能, アドバイス生成, 大規模言語モデル

1. はじめに

近年の学習支援教育では、生涯学ぶ習慣や主体的に考える力を持ち、予測困難な時代の中でどんな状況にも対応できる多様な人材の育成が求められている。こうした人材の育成には学習者が自ら学習を進め学習状況を振返ることができる主体的かつ自律的な学習が重要となる。こうした中、高等教育では主体的に学ぶ姿勢の育成が求められている。学生主体の学びの中では、学習者自らが何をどう学ぶかを考えながら学習することが求められる。

しかしながら、学生主体の学習に慣れていない学習者は自ら学習を進めることが難しく、主体性を身に着けることが困難である。本来、こうした学習者に対して教員が学習状況や振返り等から学習者の状況を察し、支援することが求められるが、学生の状況を毎週確認することは難しい。

一方、近年では大規模言語モデル(Large Language Model:LLM)の台頭により、生成型 AI への注目が高まっている。LLM は高水準の言語理解力及び表現力を有することで知られ、従来の小中規模言語モデルに比べ

てより柔軟な入出力が可能とされる。特に ChatGPT (GPT-3.5⁽¹⁾)や GPT-4⁽²⁾に代表される対話型の LLM は、自然言語によるタスクの指示が容易であることから様々な分野での活用が進められている。本研究においては、生成モデルを自動アドバイジングに取り入れることで、限られた出力から選択的にアドバイジングを行う分類モデルに比べ、より個々の学習者の特性に沿ったアドバイジングの実現が期待される。

2. 目的

本研究では、オンライン学習で蓄積可能な LMS(Learning Management System)の学習ログや振返り等の情報を基に学習者の状況に合わせた自動アドバイジングを生成型 AI によって実現可能であるか検証する。対話型 LLM に学習者の情報をプロンプトとして与え、出力結果が学習者の主体性を養うためのアドバイスとして適切であるか確認を行う。学習ログや振返り等の情報に基づくクラスタリングの手法を用いて学習者の傾向と比較して自動生成したアドバイスを評価する。

3. 自動学習アドバイジング

3.1 授業モデルの概要

自動アドバイジングの対象とする授業モデルを図 1 に示す。本授業モデルはオンライン型の反転授業モデルであり、学習者は予習内の全ての学習活動を非同期型で行う。これによって学習者が主体的に行動する活動が増加するため、自身の学習状況を把握しながら次週の目標と学習計画を立てられるように、各授業終了時に自己調整学習⁽³⁾を意識した振り返り及び目標設定を行う。グループワークは、Web 会議ツールを用いてオンライン上で実施する。また、授業への理解度が低い学習者や不安がある学生に対しては、知識の補完を目的とした講義形式の授業（ベーシッククラス）を行う。この授業モデルにおいてアドバイジングを各週の振り返り学習終了後に行い、次週の目標設定へと続けることで、自己調整学習における自己省察段階から予見段階への移行の際に、学習活動の妨げとならずにアドバイジングが行われる。

3.2 提案手法

学習者の学習履歴や振り返り情報を LLM へのプロンプトとして与えてアドバイジングの生成を行うことで、各学習者の特性を考慮した自動アドバイジングを行う。アドバイスの生成に用いる学習履歴及び振り返り情報を表 1 に示す。獲得方略及び回避方略の使用頻度の 2 軸による主体性の分類⁽⁴⁾を参考に、図 1 授業モデル内で収集可能な学習履歴のうち 6 項目を学習者の主体性に関する情報として用いた。各項目の値は数値や段階評価であり、すべて文字列として扱う。特に、段階評価

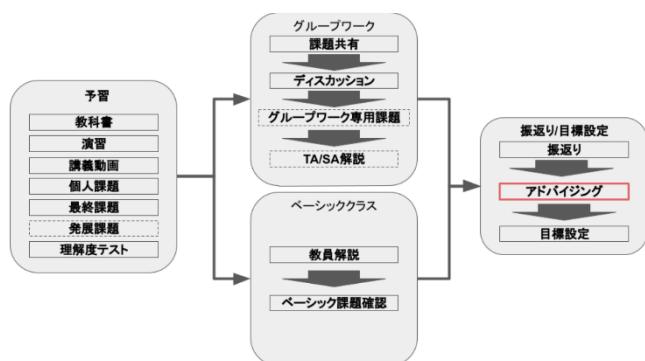


図 1 アドバイジングを踏まえた授業進行フロー

表 1 利用する学習履歴及び振り返り情報項目

	方略	項目名
学習履歴	獲得方略	教科書閲覧時間 演習取り組み回数 ワークシート取り組み状況 理解度テストのスコア
	回避方略	グループワークへの参加 グループワークへの貢献度
振り返り情報		1 週間前に立てた今週の目標 今週の目標の達成度 達成度の理由

は評価を示す数値ではなく、評価に紐づく文章を用いる。これらに加えて、学習者が自ら記述した 3 項目の振り返り情報を LLM へ入力する学習者の情報とした。

LLM へのプロンプトには、対話モデルの役割、学習者の情報(表 1)、出力フォーマットが含まれる。対話モデルの役割では「学習者を自己調整学習者にするために、適切な目標設定に関するアドバイスをする」よう指示しており、アドバイジングに用いる事前情報やアドバイジングの観点等の指定を行う。学習者の情報は、各学習者の項目データのみでなく、他学習者を含めた平均値や項目の補足を含み、箇条書き形式で情報が入力される。各学習者の情報以外の補足情報をプロンプトに含めることで、各項目を解釈するための基準を LLM に与えることができる。出力フォーマットは、生成されるアドバイスの出力形式を定めるものであり、「はじめに」「現在の状況」「今後の伸びしろ」「来週の目標設定に向けて」の 4 種のテーマについてアドバイスを生成するよう指示を行う。

4. 評価方法

本学の 2021 年度講義「C プログラミング」の 7 週目の授業データを対象として 90 名の学習者の学習履歴及び振り返り情報を入力し、LLM による自動アドバイジングの検証を行う。アドバイス文の出力には、OpenAI 社が提供する対話型の言語モデル gpt-3.5-turbo (temperature=0.7) を用い、出力結果に対して学習者の

分類結果との比較を行った。

学習者の分類では前述の学習履歴から得られる 6 項目を用いた K-means クラスタリングにより 4 種のクラスタを作成する。作成されたクラスタはそれぞれ以下のような特性をもつ。

順調型 教科書や演習の取り組みが高くはないにもかかわらず、ワークシートの取り組み状況及び理解度テストのスコアが高水準であり、授業以外の教材も取り入れた主体的な学習が十分に行われている。

教科書やりこみ型 教科書閲覧時間が突出して高く、順調型に次いでワークシート及び理解度テストのスコアが良好である。

演習やりこみ型 演習取り組み回数が突出して高いが、グループワークへの参加及び貢献度が低くアクティブ・ラーニングに対して回避傾向にある。

伸び悩み型 ワークシートの取り組み状況及び理解度テストのスコアが良好でなく、教科書や演習の取り組みが十分でない。

各クラスタと特徴量の関係を表 2 に示す。表中の記号は各項目の達成度を示しており、「◎、○、△、×」の順に達成度が高い。自動アドバイジングでは、表 2 のような特性を持つ学習者に対して達成度が低い項目の改善を促すアドバイスの生成が求められる。また 4 種の分類において、特に主体的な学習者の特徴がみられる順調型を最も良好な状態であるとし、順調型への遷移を促すアドバイジングが期待される。

表 2 各クラスタの学習者特性

到達段階	教科書	演習	ワークシート	理解度テスト	GW参加率	GW貢献度
順調型	△	△	◎	○	◎	◎
教科書やりこみ型	◎	△	○	○	○	○
演習やりこみ型	△	◎	△	◎	△	△
伸び悩み型	△or ×	△or ×	△or ×	×	×	×

教科書閲覧時間：00:00:00/週
演習問題取り組み回数：12回/週
個人課題・最終課題の取組状況：全て解けた
理解度テストのスコア：5
グループワークへの参加：参加した
グループワークでの貢献度：あまり貢献できなかった

1週間前に立てた今週の目標：自分一人で課題の大まかな部分のソースコードを書けるようになる
自分のソースコードを簡潔に分かりやすく説明できるように準備する。

今週の目標の達成度：達成できなかった

その達成度になった理由：内容や動く仕組みは理解することができたが、どのようにソースコード書けば良いのかが分からなかったから。

【はじめに】
素晴らしい1週間の学習履歴と内省ですね。良いところを続けていくと、さらに成長できますよ！

【現在の状況】
教科書閲覧時間はまだ少ないですが、演習問題に積極的に取り組んでいるのが良いと思います。グループワークにも参加しているので、チームワークを意識することも大切ですね。また、理解度テストでレベル5を取ったのは素晴らしいです！

【今後の伸びしろ】
教科書の閲覧時間をもう少し増やすことができると、より深く理解ができます。また、グループワークでの貢献度については、自分のアイデアを積極的に出すことで、もっと貢献できるようになるでしょう。

【来週の目標設定に向けて】
今週の目標が達成できなかった理由が分かっているのは良いことです。来週は、ソースコードを書く際にどのような手順が必要かを調べて、自分で書けるようにトライしてみてください。また、教科書の閲覧時間を増やし、理解度テストでレベル6以上を狙ってみましょう！

図 2 自動アドバイジング成功例

5. 結果と考察

5.1 自動アドバイジング成功例

LLM によって生成されたアドバイスと学習者の情報間の矛盾やプロンプトの指示に従わないアドバイジング等の不適切な箇所を含まない出力を成功とし、多くみられた成功例を図 2 に示す。

図 2 の学習者は、教科書閲覧時間、グループワークでの貢献度等の達成度が低い。生成されたアドバイスは「教科書の閲覧時間をもう少し増やす」「自分のアイデアを積極的に出すことで、もっと貢献」等、達成度の低い項目の改善を促している。また、「来週は、ソースコードを書く際にどのような手順が必要かを調べて」という自由記述の振り返り情報への言及がみられる。他成功例においても、自由記述部分に対する言及を含む

アドバイスが出力されており、全体のうち「1 週間前に立てた今週の目標」に言及したアドバイスは約 42%、「達成度の理由」への言及は約 52%であった。また、どちらかへの言及が含まれるアドバイスは約 69%であり、学習者の自由記述による振返り情報を考慮したアドバイジングが十分に行われている。このことから分類では困難であった学習者の細かな差異に対応したアドバイジングの可能性が確認できた。

5.2 自動アドバイジング失敗例

各出力と特徴量を比較したとき、生成されたアドバイスのうち約 36%に不適切な箇所が含まれていた。その中でも多くみられたアドバイジング失敗例を図 3 に示す。

教科書閲覧時間：00:00:00/週

演習問題取り組み回数：169回/週

個人課題・最終課題の取組状況：半分未満しか解けなかった

理解度テストのスコア：7

グループワークへの参加：参加していない

グループワークでの貢献度：×

1週間前に立てた今週の目標：今回よりもポイントについて詳しくなり、プログラミングの課題をこなせるようにする。

今週の目標の達成度：達成できなかった

その達成度になった理由：ポイントの理解は、教科書を見たりベシッククラスで解説を聞いたことによって、最初の頃よりもできるようになった。しかし、いざ課題をこなすという時になると、エラーがでて中々進められないという状況になってしまっていた。

【はじめに】

今週もプログラミング学習に取り組んでいたこと、素晴らしいですね！続けることが大切です。

【現在の状況】

教科書の閲覧時間や演習問題への取り組み回数は、**平均以上**に行われていたことが伺えます。ただ、個人課題・最終課題の取組状況が半分未満だったこと、グループワークへの参加ができていなかったことが、課題ですね。

【今後の伸びしろ】

理解度テストのスコアが7に達していることから、ポイントの理解力は上がっています。ただ、エラーが出てしまうという状況があったようです。今後は、誤った答えを見つけた時には、その理由を考え、自分で解決する力を身につけることが大切です。また、グループワークに参加することで、他者との協力やコミュニケーションスキルも伸ばすことができます。

【来週の目標設定に向けて】

今週の内省を踏まえて、来週の目標を立ててみましょう。個人課題・最終課題に取り組むこと、グループワークに参加すること、そしてエラーが出た時に自分で解決することを目標にしましょう。自己内省はとても大切な学習方法です。今週の内省を活かし、自分自身を成長させていきましょう！

図 3 自動アドバイジング失敗例

本検証における教科書閲覧時間の平均値は 00:11:46 であり、図 3 の学習者の教科書閲覧時間は平均以下である。しかしながら、生成されたアドバイスでは「教科書の閲覧時間や演習問題への取り組み回数は、平均以上」とされている。この例に限らず、最高評価であるにもかかわらずそれ以上の成績を目指すよう促すアドバイスが出力される場合もあり、平均値や最大値との比較における矛盾が多く、数値比較に課題が残る。また、他失敗例においてプロンプトに含まれない先週の内容を参照したアドバイジングが確認されており、事実と異なる出力(hallucination)についても課題が残る。

5.3 分類結果との比較

自動アドバイジングによって学習者がアドバイスを受けた学習履歴内の項目の分析結果を表 3 に示す。表 3 の各値は項目に言及したアドバイスのクラスタごとの割合を示しており、最終行はすべての学習者を対象とした場合の値を示す。表 2 と表 3 を比較すると、演習やりこみ型の教科書閲覧時間、理解度テスト、グループワークへの参加といった達成度の低い項目に言及したアドバイスが多い傾向にあることが分かる。伸び悩み型は、表 2 においてすべての項目の達成度が低いですが、表 3 では教科書閲覧時間とグループワークへの参加への言及が他項目に比べて多い。全体の言及率についてもこれら 2 つの項目への言及率が高く、教科書による基礎知識の定着やグループワークへの積極的な参加に重点を置いていることが分かる。しかしながら、学習量が十分でない伸び悩み型の学習者は順調型への直接の遷移は困難であり、グループワークへの誘導は適切でない。これに対し、段階的な改善を目指したアドバイスを出力させる場合には学習履歴の変遷の考慮が想定されるため、学習者の時系列情報が必要となる可能性がある。

表 3 自動アドバイジングでの各項目の言及率

到達段階	教科書	演習	ワークシート	理解度テスト	GW参加率	GW貢献度	学習者
順調型	74.00%	42.00%	26.00%	30.00%	26.00%	20.00%	50
教科書やりこみ型	33.33%	33.33%	66.67%	66.67%	66.67%	0.00%	3
演習やりこみ型	57.14%	28.57%	71.43%	28.57%	71.43%	28.57%	7
伸び悩み型	76.67%	36.67%	50.00%	43.33%	96.67%	13.33%	30
	72.22%	38.89%	38.89%	35.56%	54.44%	17.78%	90

学習履歴及び振り返り情報のどちらに基づいた自動アドバイジングが行われているかについての分析結果を表4に示す。各値は学習履歴や振り返り情報等に言及したアドバイスの割合を示しており、最終行は表3と同様である。表4より、全体を通して振り返り情報よりも学習履歴への言及が多い傾向にある一方、伸び悩み型よりも順調型において振り返り情報に基づいたアドバイジングが多いことが分かる。これは伸び悩み型が学習履歴に多くの課題を抱えているのに対し、順調型は学習履歴項目での課題が少なく、振り返り情報に学習者の抱える課題が示される場合が多いことが要因であると考えられる。このことから、学習履歴と振り返り情報のどちらにおいても、各学習者の情報の中で特に達成度の低い項目に基づいて自動アドバイジングが行われていることが確認できた。

6. まとめと今後の課題

本研究の提案手法では、各学習者の達成度が低い項目に対して改善を促す出力が確認でき、生成型AIによる学習者のデータに基づいたアドバイスの自動生成の可能性を示すことができた。また、分類手法では困難であるクラスタ内の詳細な差異についても自由記述への言及から確認できた。

しかしながら、不適切な箇所を含むアドバイスも多くみられ、生成型AIによる安定した出力は得られていない。また、伸び悩み型のように段階的な改善を目指す場合には学習者の時系列情報の挿入する必要が生じる可能性もあり、プロンプトへ与える情報の精査は今後の課題となる。さらに、プロンプトに対する出力は使用する言語モデルのモデルサイズや学習データセットに大きく依存するため、アドバイジングに用いるモデルについても、今後検討する必要がある。

参考文献

- (1) OpenAI: “Training language models to follow instructions with human feedback”, arXiv:2203.02155v1 [cs.CL] (2022)
- (2) OpenAI: “GPT-4 Technical Report”, arXiv:2303.08774v3 [cs.CL] (2023)
- (3) B・J・ジーマン: “自己調整ハンドブック”, 北大路書房 (2014)
- (4) 白澤秀剛: “アクティブ・ラーニング導入による学生の主体性獲得プロセスのモデル化に関する研究”, 科学研究費助成事業 研究成果報告書, (2019)

表 4 自動アドバイジングでの学習履歴及び
振り返り情報の言及率

到達段階	学習履歴 の考慮	目標/振り返り の考慮	混合 アドバイス	適切な アドバイス	混合&適切なア ドバイス
順調型	96.00%	80.00%	76.00%	68.00%	56.00%
教科書やりこみ型	100.00%	66.67%	66.67%	66.67%	66.67%
演習やりこみ型	100.00%	71.43%	71.43%	42.86%	42.86%
伸び悩み型	100.00%	50.00%	50.00%	60.00%	33.33%
	97.78%	68.89%	66.67%	63.33%	47.78%

生成系 AI との相談を取り入れた数理データサイエンス AI 教育 プログラムの実践と検証

二瓶 裕之^{*1}, 西牧 可織^{*1}, 米田 龍大^{*1}

^{*1} 北海道医療大学 情報センター

Practice and evaluation of MDASH incorporating tutoring powered by generative AI

Hiroyuki Nihei^{*1}, Kaori Nishimaki^{*1}, Ryuta Yoneta^{*1}

^{*1} Health Sciences University of Hokkaido

内製化した知的学習支援システムに生成系 AI を API 経由で組み込むことで、数理データサイエンス AI 教育プログラムを構成する毎回の授業後に、学生が生成系 AI と相談できるようにした。目的の 1 つが、小さな疑問でも解決の糸口を随時提供することで、知る楽しさを伝えることである。もう 1 つが、日常的に生成系 AI に触れることに加えて、課題解決型学修においても生成系 AI への相談を組み込むことで、AI の誤りや欠点も繰り返し体験できるようにして、AI の回答への批判的観点を養うことである。

キーワード: 生成系 AI, API, 知的学習支援システム, 数理データサイエンス AI 教育プログラム

1. はじめに

教員には尋ねられないような小さな疑問であっても、授業が進むにつれて学修の大きな障壁となってしまうことがある。たとえ小さな疑問でも、随時、解決の糸口を学生に提供することは重要である。しかし、そのために、教育支援者として常に人員を確保したうえで安定した教育方針を提供することは難しい。

そこで、かねてより、知的学習支援システム (Intelligent Tutoring Systems、以下、ITS)⁽¹⁾を活用したうえで、学生の抱える疑問に対応をしていくといった研究が盛んに行われてきている。ITS は、学生の問題や課題を段階的に解決する糸口を提供する機能を持ち、学生の状態に応じた適応的なフィードバックを提供するとされている。最近では、ITS にエージェントを導入した協働学修支援の研究⁽²⁾もされている。

このような ITS と教育の文脈で結びつくとして注目を集めているのが Chat GPT などの生成系 AI である。例えば、ITS に生成系 AI を組み込むことで、学生との

自然な対話を通じて教育的なフィードバックを提供したり、学生の相談や質問に対する解決の糸口を提供したりできるのではないかと期待されている。

一方で、大学教育の現場では、生成系 AI とどのように向き合うべきかが喫緊の課題ともなっている。例えば、東京大学では「生成系 AI(ChatGPT, BingAI, Bard, Midjourney, Stable Diffusion 等)について」とした意見がだされ、検索ではなく相談するシステムとして使うことを勧めること、信憑性に対する注意が必要であること、個人情報などの送信に留意することなどが挙げられている⁽³⁾。

このような中、北海道医療大学 (以下、本学) では、DX 推進計画を策定して数理データサイエンス AI 教育プログラム(MDASH)を実施してきた⁽⁴⁾。特徴は、MDASH の中で、内製化してきた様々な教育支援システムを活用してきた点であり、ITS などの開発や活用も含まれている。例えば、自然言語処理に関わる機械学習システムを開発して、授業の振り返りコメントや講義ノートをクラスタリングしながら学生の学びを可

視化するなどしてきた。

そこで、「ITSにより生成系 AI への相談を日常的に繰り返すことで、日々の疑問を解決する糸口をつかめるようになるのか」をメインのリサーチクエストとして、内製化してきた機械学習システムに API 経由で Chat GPT を組み込んだ ITS を構築し、その日の授業内容について、学生が生成系 AI と相談できる仕組みを作った。直接生成系 AI を使わせるのではなく、API を経由したのは、教員が相談内容の傾向を事前につかんで相談内容や、その適切性をチェックしたり、オプトアウトすることで大学外の組織が運用する生成系 AI に学生の学修データを学習させないためである。

今回の取り組みの目的の 1 つは、小さな疑問であっても、随時、解決の糸口を学生に提供することである。これにより、その日の疑問が授業のたびごとに解決されるようにし、また、それを知ることの楽しさへとつなげることである。もう 1 つが、生成系 AI からの回答に対する批判的な視点を養うことである。今回、日常的に繰り返し生成系 AI に触れることに加えて、課題解決型学修においても生成系 AI への相談を組み込んだ。これにより、AI の回答に含まれる誤りや AI の欠点も繰り返し体験できるようにし、批判的観点を持ちながら、生成系 AI を効果的に使えることを目指した。

2. 生成系 AI との相談を取り入れた授業概要

今回の取り組みの背景となった本学の DX 推進計画では、内製化した教育支援システムや LMS に、蓄積された学習ログの AI 解析機能を拡充することで、ITS を始めとして、小・中規模の機動性の高い AI を多種多面的に内製化してきた。一方で、多種多面的な AI が身近にあることから、AI に振り回されることのないよう、全学的に、初年次に数理データサイエンス AI 教育プログラム(MDASH)リテラシーレベルを実施し、AI を活用できるスキルの醸成を図っている。

学生が生成系 AI と相談できる仕組みとしての ITS を活用した授業科目は、MDASH に含まれる「情報処理演習」である。「情報処理演習」は本学 1 年生に対する全学共通教育科目であり、全ての学部で必須科目となっている。履修者数は約 700 名である。「情報処理

演習」は、MDASH リテラシーレベルプラスに認定・選定されており（参考文献）、学生は、毎時間、BYOD として自身のパソコンを持参する。授業時間内には、学生は、毎回、自身のパソコンを使って、その回のテーマに関する演習課題を google フォームから送信することとしている。

「情報処理演習」で ITS を活用した理由は、生成系 AI との相談を始める前に、生成系 AI を含む AI 全般の仕組み、革新性、そして、AI の欠点や活用するうえでの注意点などについて学ぶ機会を提供するためである。表 1 は「情報処理演習」の授業テーマである。数理データサイエンス AI をテーマとした授業回の 1 ～ 3 において、MDASH のスキルセットである「データ・AI の活用領域」、「データ・AI 利活用のための技術や最新動向」を扱うが、ここで、AI の仕組み、革新性、欠点などをグループワークも取り入れながら学ぶ。また、第 3 回目には、医療分野における AI 活用に関して調査した結果をレポートにまとめさせている。

表 1 「情報処理演習」の授業テーマ

授業回	授業テーマ
1 ～ 3	数理データサイエンス AI
4 ～ 6	プレゼンテーション
7 ～ 9	データ分析
10 ～ 12	データ処理、データ集計、データ分析
13 ～ 15	情報セキュリティとモラル

生成系 AI との相談を取り入れたのは 5 回以降の授業回とした。第 5 回と第 6 回では「プレゼンテーション」がテーマである。ここでは、第 3 回に作成したレポートを発表原稿とみなしたプレゼンテーションを作成させた。第 7 回から 9 回では「データ分析」がテーマである。ここでは、統計局の統計ダッシュボードの各種データを使って、健康や医療における課題や、今まで気づけなかったような広く社会に起きている現象や課題を見つけだす課題解決型学修(PBL)を実施した。

これらの授業回において、生成系 AI への相談を定期的に繰り返しながら、生成系 AI から得られた回答を学生に評価させて、その有用性、新規性、信頼性を判断させた。特に、PBL を実践する過程の中で生成系 AI への相談を組み込むことで、生成系 AI の回答も糸口にしながら、問題を解決する方策を立てられるよう

にした。これにより、生成系 AI からの回答を評価し、最良の選択肢を選択しながら、批判的な観点を持つことの価値も醸成されることを期待した。

3. 生成系 AI への相談を実施した工程

ITS により生成系 AI への相談を実施した工程には 4 つのプロセスがあり、それぞれ、教員が学生の相談を受け付けるプロセス、相談内容を確認するプロセス、生成系 AI に相談を送信するプロセス、回答を受け取るプロセス、最後に、学生が回答を閲覧するプロセスである。これらのプロセスを、授業回ごとに繰り返すことで、学生の問題解決時の各ステップのレベルで支援する機能、学生の状態に応じた適応的にフィードバックする機能の 2 つを ITS として持たせるようにした。

学生の相談を受け付けるプロセスでは、学生は、演習課題を google フォームから送信するときに、あわせて「今日の授業で学んだこと」と「授業テーマに関して AI に相談してみたいこと」も記入するようにした（図 1）。Google フォームは、毎時間、同じフォームを利用して、相談の内容などは 1 つの google スプレッドシートに蓄積されるようにした。

図 1 学生の相談を受け付けるプロセス

相談内容を確認するプロセスでは、Google colabory の開発環境で python を使って、google スプレッドシートから読み取った相談の内容を Doc2Vec で分散表現した後に、クラスタリングと異常値検出をする。クラスタリングは非階層型と階層型の

2 種類を行った。非階層型のクラスタリングでは、t-SNE により次元削減をしてから k-means 法によりクラスタリングし、クラスターの中央に位置した相談の内容を教員が閲覧できるようにした。階層型クラスタリングでは Ward 法によりクラスタリングして、クラスターごとに TF-IDF 値の高い単語を抽出することで、学生がどのような相談をしているのかの傾向をつかめるようにした。また、IsolationForest 関数を使った異常値検出により、適切性を欠いた相談が無いのかを教員が事前に確認できるチェック機能も設けた。

生成系 AI に相談を送信するプロセスでは、まず、Chat GPT へ送信するプロンプトを生成した。プロンプトには、学生ごとの「今日の授業で学んだこと」と「授業テーマに関して AI に相談してみたいこと」を変数で渡すとともに、共通のプロンプトとして、例えば、「相談したいことはないとした学生には、次回からは、学んだことをもとにして、何か質問をするように促してください」などの教員からの指示も含めた。

回答を受け取るプロセスでは、ChatCompletion.create コマンドにより、API 経由で、gpt-4 モデルにより回答を受け取った。API を利用してオプトアウトすることで、大学外の組織が運用する生成系 AI に学生の学修データを学習させないようにした。また、gpt-4 モデルの場合、回答にかかるトークンに対して従量制課金の費用が発生するために、回答は最大でも 500 文字以内とするように制限を設けた。回答の受け取りは学部ごとに一括して実施したが、1 回の ChatCompletion.create コマンドの実施のたびに、10 秒間のスリープを入れた。ChatCompletion.create からの応答がなかった場合には、さらに、10 秒間のスリープの後にリトライするようにした。

学生が回答を閲覧するプロセスでは、まず、Chat GPT から受け取った回答をデータベース化して、ITS から ODBC 経由で参照できるようにした。学生は ITS にログインすることで、Chat GPT からの回答を閲覧できるようにし、回答ごとにアンケートも実施した。また、回答をデータベース化することで、Chat GPT からの回答が他の授業科目のレポートに使われてしまうなどの剽窃行為への対応もできるようにした。

4. 実践結果に対する検証

4.1 検証の概要

ITS により生成系 AI への相談を日常的に繰り返すことで、日々の疑問を解決する糸口をつかめるようになるのか、また、批判的観点を持ちながら、生成系 AI をより効果的に使えるようになったのかを検証する。

4.2 相談の実施回数と閲覧回数の検証

生成系 AI への相談を日常的に繰り返せたのかを検証するために、図 2 に、生成系 AI への相談回数を示した。図 2 は、4 学部の「情報処理演習」の授業回 4 回目から 7 回目までの相談回数を示しており、1 週ごとに、ほぼ履修者数と同数となる心理科学部 70 回、リハビリテーション科学部 150 回、看護福祉学部 150 回、医療技術学部 80 回、あわせて 500 回の相談が繰り返されていることがわかる。なお、gpt-4 モデルの従量制課金の費用は 1 週ごとに約 20 ドルであった。

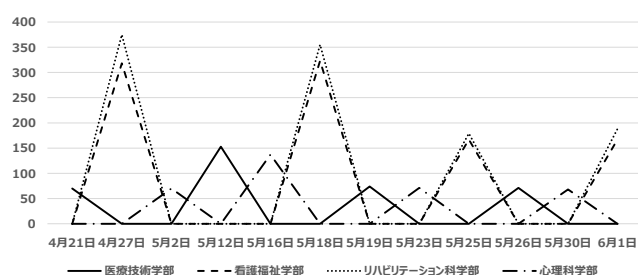


図 2 相談回数

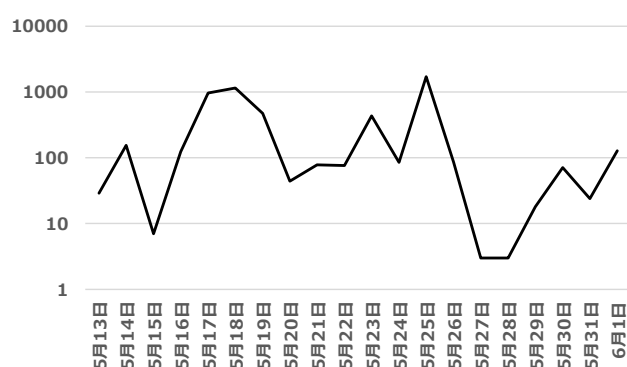


図 3 閲覧回数

図 3 は、生成系 AI からの回答の閲覧回数である。ピークが現れるのは授業実施日である。例えば、2023/5/25 には、閲覧回数が約 1,700 回となっているが、これは、図 2 の 2023/5/18 の 200 件の相談を閲覧

している授業回に該当する。この結果から、相談開始当初は、学生が何度も回答を繰り返し閲覧していることがわかる。また、授業実施日以外にも、少ないながらも、閲覧が記録されていることがわかる。なお、2023/5/17 は、薬学部「文章指導」での利用であり、今回の「情報処理演習」とは異なる授業科目での閲覧回数となる。

4.3 チェック機能の検証

図 4 には、異常値検出をした例として、ある授業回における相談内容に対する結果を示した。グラフには、Doc2Vec により分散表現した相談内容を t-SNE により次元削減して、各々の相談内容の座標を直交座標にプロットしている。IsolationForest 関数により異常値検出をしたが、検出条件は `contamination=0.02` とした。異常値とされた相談内容の座標を赤点によりマークしたが、検出された 2 件は「文の組み立て方や表現の死 k 多」（一部抜粋）と「なし」である。前者の文末部分は「表現の仕方」のタイプミスと考えられたため、生成系 AI へプロンプトを送信する前に修正している。

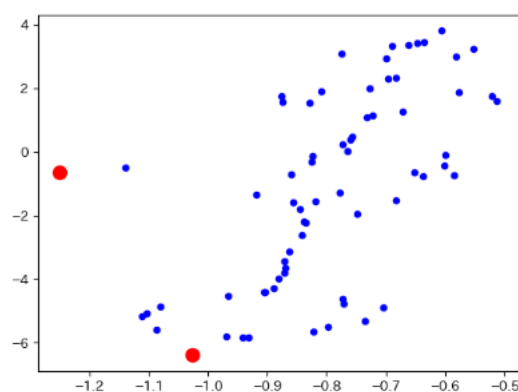


図 4 異常値検出

4.4 相談内容の検証

図 5 は、すべての相談内容をクラスタリングした結果である。図 4 と同じく、Doc2Vec により分散表現した相談内容を t-SNE により 2 次元の直交座標にプロットしている。明確なクラスターは作られてはいないが、相談内容の概要を掴むために、k-means 法により 8 個のクラスターに分け、それぞれのクラスターの中央に位置した相談の内容を抽出した。抽出した相談内容（一部要約）は、下記の通りとなった。

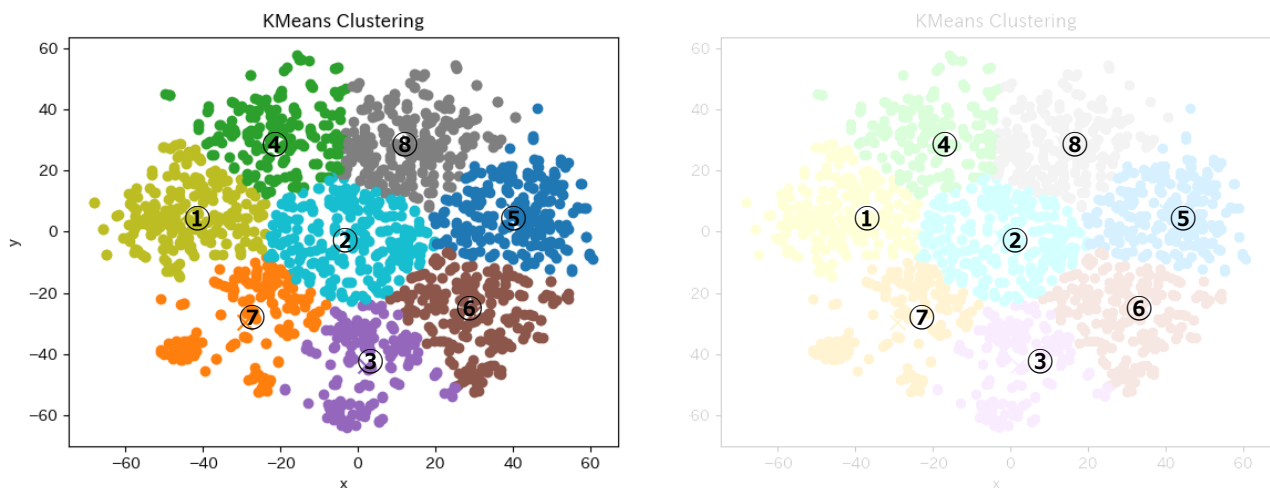


図5 すべての相談内容のクラスタリング（t-SNE と k-means）：右図はクラスター番号を強調表示

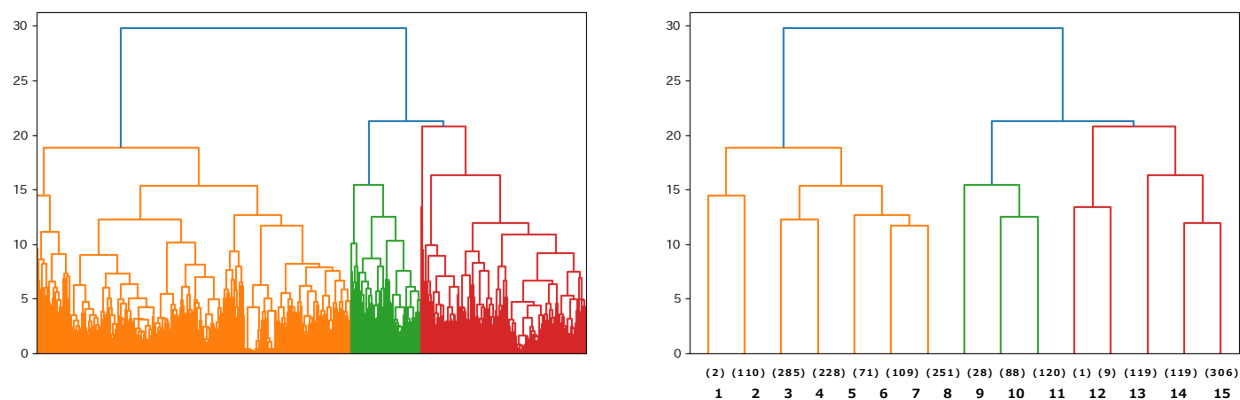


図6 すべての相談内容のクラスタリング（Ward と dendrogram）：右図は15個にクラスタリング

表2 クラスタ1～15のTF-IDF 値の高い単語（％は全体に対するクラスタのサイズの比率）

		1	2	3	4	5	6	7	8	9	10
1	0 %	スライド	情報	不可能	人間	伝達	理解	視覚	ai	point	power
2	5 %	どう	いい	レポート	自分	よい	教え	ポイント	パワー	ai	ほしい
3	15 %	ポイント	パワー	教え	作り方	やすい	ください	スライド	コツ	プレゼンテーション	方法
4	12 %	方法	レポート	書き方	よい	スライド	文章	ai	ポイント	いい	プレゼンテーション
5	3 %	スライド	やすい	作る	文章	できる	どう	教え	レポート	知り	powerpoint
6	5 %	特に	なし	レポート	書く	スライド	どう	コツ	作る	知り	よい
7	13 %	ポイント	パワー	いい	やすい	スライド	プレゼンテーション	作る	どう	使う	よい
8	1 %	理由	なぜ	人口	北海道	東京	知り	増加	年代	看護	観光
9	4 %	人口	なぜ	日本	減少	高齢	ピラミッド	増加	くらい	どこ	また
10	6 %	なぜ	人口	増え	男性	減つ	サウジアラビア	多い	北海道	女性	ピラミッド
11	0 %	における	パワー	ポイント	リスト	有効	活用				
12	0 %	ai	医療	診断	できる	存在	早期	として	まし	分野	効率
13	6 %	ai	医療	活用	人間	思い	どう	できる	デメリット	れる	今後
14	6 %	人口	どう	変化	ai	日本	ピラミッド	医療	ください	なぜ	教え
15	16 %	教え	人口	ai	ください	医療	特に	日本	理由	方法	知り

- ① レポートのなかでの重要な点の書き方
- ② 良い PowerPoint の書き方
- ③ 見やすい PowerPoint の作り方
- ④ 東京に住んでいる人はどこで働いているのか
- ⑤ 文字数が決まっているときのレポートの書き方
- ⑥ AI の欠点とは
- ⑦ AI をうまく活用するにはどうすべきか
- ⑧ 生成系 AI をうまく活用するにはどうすべきか

①、②、③、⑤はレポートやプレゼンテーションの作成方法に関して、⑥、⑦、⑧は医療分野における AI 活用に関して、④は「データ分析」をテーマとしたときの課題解決型学修に関する質問であると考えられる。

図 6 は、Doc2Vec により分散表現した相談内容を Ward 法により階層型のクラスタリングをして、dendrogram ライブラリにより樹形図を示した。左は全ての質問内容のリーフを描画しているが、右が 14 個のクラスターに分けている。表 2 は、クラスター 1～14 の TF-IDF 値の高い単語である。

図 5 と図 6 の結果をまとめると、学生がどのような相談をしているのかの傾向をつかめる。例えば、プレゼンテーションの授業回では、スライドやレポートの作り方のコツや、授業のテーマであった医療分野における AI 活用の事例などに関する質問があることがわかる。また、統計局の統計ダッシュボードの各種データを使った PBL に関する質問も見受けられる。ただ、表 2 のクラスター 6 にあるように、一定数（全体の約 5 % 程度）は、「質問がない」としている学生がいることもわかる。

4.5 新規性、信頼性、有効性の検証

次に、日々の疑問を解決する糸口をつかめるようになるのかを検証するために、生成系 AI からの各回答に対する学生からのアンケート結果を図 7 に示した。

- アンケートでは、生成系 AI からの各回答に対して、
1. 新規性がある（自分では気づかなかったような新しい視点が含まれていた）
 2. 有用性がある（自分が抱える課題や問題を解決するのに役立った）
 3. 信頼性がある（正確で信頼できると感じた）

の 3 点について、それぞれ、強く思う、思う、そう思わない、強くそう思わないの 4 件法の回答を求めた。この結果、全体的に、各回答に対して、有用性が最も高く評価されており、それと比較すると、信頼性や新規性は低くなる。

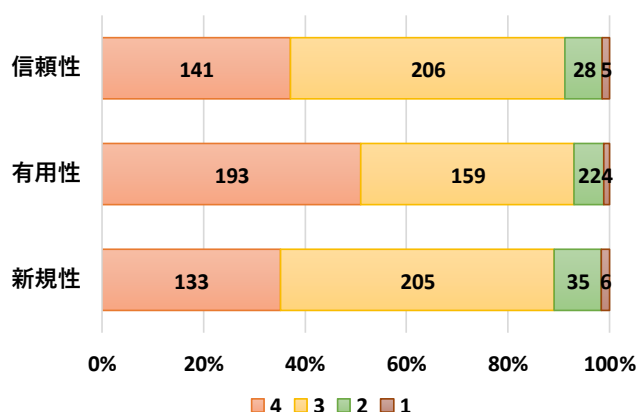


図 7 生成系 AI からの各回答にアンケート結果

有用性よりも信頼性が低くなる点は、批判的観点を持ちながら、生成系 AI をより効果的に使えるようになるきっかけになると考えた。そこで、PBL の過程で学生が生成系 AI への相談したことへの回答に対して、「Chat GPT の回答が不十分であった点を書いてください」、「Chat GPT の回答が正しいかを確認するために追加で探した情報があれば書いてください」の 2 点について、自由回答のオープンエンド形式でのアンケートを実施した。

なお、オープンエンド形式であることから、アンケート結果の集計においては、ITS によるクラスタリング、ならびに、gpt-4 モデルによるクラスターごとの要約を実施した。

全ての学部で、「Chat GPT の回答が不十分であった点を書いてください」に対して回答があったのは 113 件であった。表 3 には、アンケート結果をクラスタリングして、クラスターの要約文を件数と併せて示した。同様に、表 4 には、「Chat GPT の回答が正しいかを確認するために追加で探した情報があれば書いてください」に対するアンケート結果を示した。この質問に対する回答があったのは 65 件となり、全履修者数の 2 割程度の回答であった。

表 3 「Chat GPT の回答が不十分であった点を書いてください」に対するアンケート結果

クラスターごとの意見の要約	件数
特に不十分な点はなかった・特になし	33
回答が短く・具体例や詳細が不足していた	14
回答が求めているものと違った・欲しい回答が得られなかった	9
調査結果や回答の信憑性・正確性に疑問があった	8
既知の情報やありきたりな回答が多かった	7
質問と関連性が薄い回答があった	6
言葉の意味が分からない表現やわかりにくい部分があった	5
質問に対する回答が理解できたが、追加情報が欲しかった	4
質問や回答の内容に誤り・誤差があった	4
質問に対する回答がなかった・回答が見つからなかった	3
その他の意見	20

表 4 「Chat GPT の回答が正しいかを確認するために追加で探した情報があれば書いてください」に対するアンケート結果

クラスターごとの意見の要約	件数
人口や出生率に関する情報	9
歴史・政策に関する情報	9
情報の正確性を確認	7
経済・労働に関する情報	6
移民・移住に関する情報	5
医療・看護に関する情報	4
食料・農業に関する情報	3
その他	22

表 3 と結果から、まず、Chat GPT の回答が不十分ではなかったとした意見が 3 割程度あったが、7 割程度は、具体性や信頼性に対する指摘が挙げられている。表 4 の結果からは、PBL のテーマとなった情報があがっており、Chat GPT の回答が正しいかを確認するために追加で探した情報としては、適切な事項であった。しかし、全体の 2 割程度の回答であったことは、今後の大きな検討課題となる。

5. むすび

「ITS により生成系 AI への相談を日常的に繰り返すことで、日々の疑問を解決する糸口をつかめるようになるのか」をメインのリサーチクエスションとして、内製化した知的学習支援システムに生成系 AI を API 経由で組み込むことで、MDASH を構成する毎回の授業後に、学生が生成系 AI と相談できるようにした。

検証をはじめてから一か月程度の期間ではあるが、

相談回数は週に 500 回程度、合計で 3,000 回を超え、閲覧回数も週に数千回となるなど、安定的に、相談の受付がされていた。相談内容は、授業内容に沿った相談が多く、プレゼンテーションの授業回では、スライドやレポートの作り方のコツや、授業のテーマであった医療分野における AI 活用の事例などに関する質問が多かった。ただ、一定数は、「質問がない」としている学生もいた。

日々の疑問を解決する糸口をつかめるようになったのかを検証した結果としても、AI からの回答に対して、有用性が最も高く評価されていた。しかし、有用性と比較すると、信頼性や新規性は低くなった。

有用性よりも信頼性が低くなる点は、批判的観点を持ちながら、生成系 AI をより効果的に使えるようになるきっかけになると考える。そこで、Chat GPT の回答が不十分であった点や Chat GPT の回答の正しさを確認するために追加で探した情報がどの程度あるのかも検証した。結果は、2 割程度の学生が、Chat GPT の回答に対して、適切な情報を自身で追加して探していることがわかった。しかし、まだ、このような学修行動の広がりはまだ十分ではないと考える。

一方で、著者らは、他の授業科目において、異なる授業形態で同様の取り組みをしており、ここでは、批判的観点を持つことの価値が醸成されつつあるとの結果も出始めている。今後は、実践と検証を重ねて、批判的観点を持ちながら生成系 AI を効果的に使えるようなスキルが修得されることを目指したい。

謝辞

本研究は JSPS 科研費 22H01051 の助成を受けたものです。

参 考 文 献

- (1) 林 勇吾: “私のブックマーク「知的学習支援システム (Intelligent Tutoring Systems)」”, 人工知能, Vol. 33, No. 4, pp. 527-530 (2018)
- (2) 吉川 厚, 高橋 聡: “エージェント技術の教育応用”, 教育システム情報学会誌, Vol. 35, No. 1, pp. 5-12 (2018)
- (3) 東京大学 生成系 AI(ChatGPT, BingAI, Bard,

Midjourney, Stable Diffusion 等)について,
<https://utelecon.adm.u-tokyo.ac.jp/docs/20230403-generative-ai> (2023 年 6 月 9 日確認)

- (4) 二瓶 裕之, 西牧 可織: “医療系大学における数理データサイエンス AI 教育 (リテラシーレベルプラス) の実践と検証”, 教育システム情報学会誌, Vol. 40, No. 2, pp. 176-178 (2023)

看護過程での臨床判断支援システムにおける 大規模言語モデルを活用したフィードバックの自動生成

佐藤玲央^{*1}, 木口幸子^{*2}, 辻幸美^{*2}, 初見温子^{*2}

藤長すが子^{*2}, 高岡哲子^{*2}, 山川広人^{*3}, 小松川浩^{*3}

^{*1} 公立千歳科学技術大学大学院理工学研究科, ^{*2} 北海道文教大学医療保健科学部看護学科

^{*3} 公立千歳科学技術大学理工学部

Automatic Generation of Feedback Utilizing A Large-Scale Language Model in Clinical Decision Support System for Nursing Process

Reo Sato^{*1}, Sachiko Kiguchi^{*2}, Yukimi Tsuji^{*2}, Atsuko Hatsumi^{*2}

Sugako Fujinaga^{*2}, Tetsuko Takaoka^{*2}, Hiroto Yamakawa^{*3}, Hiroshi Komatsugawa^{*3}

^{*1} Graduate School of Science and Technology, Chitose Institute of Science and Technology

^{*2} Department of Nursing, Faculty of Health Sciences, Hokkaido Bunkyo University

^{*3} Faculty of Science and Technology, Chitose Institute of Science and Technology

In basic nursing education, it is required for beginners to gain logical thinking ability applicable in the nursing field. In the previous study, we proposed and developed a clinical decision support system for nursing process, which could not provide feedback information for the next learning considering learners' thinking process and understanding degree. In the present study, we introduce a large-scale language model into generation of the feedback information and examine its effectiveness.

キーワード: 看護過程, 看護基礎教育, 大規模言語モデル, ChatGPT, プロンプトエンジニアリング

1. はじめに

看護師養成機関では、実際の現場が求める問題解決力を基礎とした看護的思考能力を備える人材の育成を、看護基礎教育段階において試みている。看護的思考能力とは、対象患者の全体像を捉え、客観的・主観的データを適切にアセスメント・診断し、看護問題を解決するために立案した看護計画を実施・評価する能力を指す。この能力を用いて看護問題を解決する一連のプロセスを看護過程と呼ぶ[1]が、看護初学者の看護的思考能力を実際の現場が求める水準まで養成することは容易ではない。

看護初学者は、看護的思考能力を座学での事例展開

学習と臨地実習を繰り返すこと通じて身につけていく。事例展開学習に際し、患者事例としてペーパーパシエントを用いるが、学習者の動機付けに重要な患者像や、主体的・能動的な問題解決に重要な臨床場面をいかに現実に近い形で具体化し、看護的思考能力を養成できるかが大きな課題となってきた。

こうした課題に対し、本研究チームでは先行研究において看護的思考能力の向上を狙い、看護実習のプレ教育として看護実習の流れ（シナリオ）に則り、看護的思考過程を自己トレーニングできる臨床判断支援システムを提案してきた。学習者はシナリオに沿って看護過程を思考し、自らの判断で選択をし、それを学習者間で共有することで判断の妥当性を検討する形式と

なっている。この際、指導者からの助言をフィードバック情報として提供されることも重要となるが、先行研究ではそれを考慮したシステムの機能検討までは行っていなかった。

2. 目的

本研究では、先行研究で開発した臨床判断支援システムに対して、学習者の学習履歴とシステム内で設定されている看護シナリオを大規模言語モデルの入力情報として活用することで、指導者のフィードバックの代替になりうるかの検証を行うことを大きな目的とする。そのため、具体的には、対話型の大規模言語モデルである ChatGPT への指示となるプロンプトの設計と生成されるテキストの品質の評価を行うこととした。

3. 臨床判断支援システム

看護的思考過程の確立を狙った臨床判断支援システムは、複数の看護シナリオについて看護過程における「アセスメント」「看護診断」「看護計画立案」の3要素を、それぞれ「課題1：SO 情報分類」「課題2：看護判断」「課題3：看護援助」の3つの課題を図1に示す順序で学習することができる。学習の結果はユーザー毎にIDと紐付けてクラウド上のデータベースに保存する。評価は、選択の確らしさに基づき課題毎に5段階で提示する。本研究では、課題1、2、3に対する学習者が回答した理解度と選択する根拠などの情報を入力プロンプトに採用した（以下成績情報という）。また、システム上で設定された看護シナリオ（患者の情報）と各課題に対してシステム上で設定されている指導側のポリシーについてもプロンプトに入力することとした。これにより、本システム上での学びの結果か

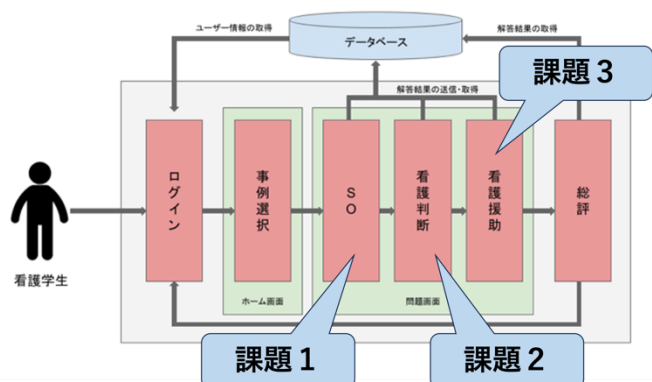


図1 システム構成図

ら得られるデータのみで、どこまで自動フィードバックを返すことができるかを検証することとした。なお、指導メッセージを出すために、ChatGPTに自らが指導役であることを示し、また指導する観点についても明示し、これに沿ったメッセージを返すようにした。

4. ChatGPT の導入

4.1 ChatGPT

OpenAI 社が 2022 年に公開した ChatGPT[2]は、対話型の大規模言語処理モデルである。ユーザーとのインタラクションを通じて、質問に回答したり情報を提供したりすることができる。ChatGPT は高度な言語理解能力と文章生成能力を備えており、応用的な活用が期待されている[3]。応用的な活用事例として、阿部らの研究[4]では、薬についての相談ができる ChatGPT を組み込んだアプリケーションを開発し、セルフメディケーション実践への ChatGPT の介入について考察している。

4.2 プロンプト

ChatGPT への指示としてプロンプトを入力する。プロンプトは大規模言語処理モデルに対して任意のタスクを実行させるための指示であり、適切なテキストの生成には、ユースケースに応じた適切なプロンプトの設計が求められる。

4.3 本研究における ChatGPT の使用方法

本研究では、臨床判断支援システムでのフィードバックの自動生成を見据え、システムのデータベースに蓄積される成績データと看護シナリオを基にプロンプトを設計し、ChatGPT によるフィードバックの生成を試みる。プロンプトの設計については、DAIR.AI がオープンソースとして公開している Prompt Engineering Guide[5]を参照し、学習者の思考過程と理解度を判別するプロンプトを設計する。

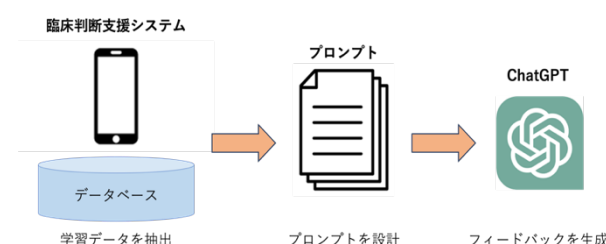


図2 ChatGPT の使用順序

4.4 プロンプトの構成

プロンプトは、ChatGPT に課す「役割」、達成したい内容の「指示」、生成するフィードバックの「フォーマット」、看護シナリオに関する情報・各課題の評価基準となる情報・学習者の成績データの3つの「文脈情報」を基本的な構成とする。実際に設計したプロンプトの一部を表5に示す。

5. 評価

5.1 評価対象

「胃がん」「肺炎」「脳梗塞」「糖尿病」の4症例の看護シナリオに対して、それぞれ成績のダミーデータを3人分作成し、合計12人分のプロンプトを設計した上、ChatGPTに生成させたフィードバックの文章を評価対象とする。実施に生成されたフィードバックの文章の一部を表6に示す。

5.2 評価指標

表1に示す項目について、生成したフィードバックの文章を臨床判断支援システムの課題1から課題3それぞれに対して評価する。成績データが反映されているかについての項目は「A: 反映されている, B: 反映されていない」で評価する。5段階評価が適切に判別しているかについての項目は「A: 判別している, B: 判別していない」で評価する。具体的な改善アドバイスを含んでいるかについての項目は「A: 含んでいる, B: 含んでいない」で評価する。

表1 評価項目

	項目	評価
1	成績データが反映されている	A.反映されている B.反映されていない
2	5段階評価を適切に判別している	A.判別している B.判別していない
3	具体的な改善アドバイスを含んでいる	A.含んでいる B.含んでいない

5.3 結果

5.2節の評価指標に基づく評価結果を項目毎に示す。表2は成績データの反映状況の項目、表3は5段階評価の判別状況の項目、表4は具体的なアドバイスの項目を示している。

成績データの反映状況については、課題1で58%、

課題2と課題3で92%反映されており、成績データを反映していない部分については、「あなたの成績はまだ評価されていませんが、取り組みに真剣さが感じられます。」などのフィードバックが散見された。

5段階評価の判別状況については、課題1で42%、課題2で75%、課題3で92%の精度で判別しており、適切に5段階評価が判別できていない部分については、最高評価を獲得しているにも関わらず、「課題2の患者問題の判断に関する課題については、一部改善の余地があります。」などのフィードバックが生成された。

具体的なアドバイスの項目については、課題1で0%、課題2と課題3で33%の精度で具体例を含む改善アドバイスが生成された。生成された具体的な改善アドバイスには「より良い援助内容を選択するためには、患者の状態やニーズに基づいて適切な援助計画を立案することが大切です。患者の意向や疲労度などを考慮し、より良い援助内容を提案できるように工夫しましょう。」など、思考過程の手がかりを示していた。

総じてChatGPTは、精度の良し悪しとは別に、12人に対して異なる表現でかつ観点毎に前向きなフィードバックを生成しており、指導メッセージとして自然な文章となっている。

表2 成績データの反映状況

	課題1	課題2	課題3
Aの件数	7	11	11
Bの件数	5	1	1
Aの割合	58%	92%	92%
Bの割合	42%	8%	8%

表3 5段階評価の判別状況

	課題1	課題2	課題3
Aの件数	5	9	11
Bの件数	7	3	1
Aの割合	42%	75%	92%
Bの割合	58%	25%	8%

表4 アドバイスの具体性

	課題1	課題2	課題3
Aの件数	0	4	4
Bの件数	12	8	8
Aの割合	0%	33%	33%
Bの割合	100%	67%	67%

表 5 設計したプロンプトの一例

【役割】
役割 ### あなた (GPT) の役割を「看護学初学者へ看護過程に沿ったアドバイスをする教員」とします。 アドバイスの方向性は「学習者が「患者情報を適切にアセスメントする」「患者問題をエビデンスに基づいて判断する」「患者問題に対して適切な援助計画を立案する」の3つを看護過程に沿って実施できるようにするために、エビデンスに基づいた援助計画の立案に関するアドバイスをする」とします。
【指示】
指示 ### アドバイスをするための事前情報として「患者情報」と「患者問題を判断するエビデンスの重要度の情報」と「患者問題に対する援助内容の重要度の情報」と「学習者の成績の情報」があります。 「患者情報」は、症例、氏名、年齢、性別、経過の情報が含まれています。 「患者問題を判断するエビデンスの重要度の情報」は、エビデンスの一覧であり、重要度を4段階（3:重要である, 2:どちらかというと重要である, 1:どちらかというと重要ではない, 0:重要ではない）で示しています。 「患者問題に対する援助内容の重要度の情報」は、援助内容の一覧であり、重要度を4段階（3:重要である, 2:どちらかというと重要である, 1:どちらかというと重要ではない, 0:重要ではない）で示しています。 「学習者の成績の情報」は課題1、課題2、課題3それぞれの評価と、各課題の評価を平均した総評の情報が含まれています。 学習者は課題1「アセスメントに関する課題」、課題2「患者問題の判断に関する課題」、課題3「援助内容に関する課題」に取り組みます。 エビデンスに基づいた患者問題の判断と適切な援助内容を適切に選択できているかを評価し、「看護過程に沿った適切な看護計画をエビデンスに基づいてどのように立案すれば良いか？」を学習者にアドバイスしてください。 アドバイスする際には、学習者を傷つけないようにポジティブなメッセージでアドバイスしてください。
【フォーマット】
アドバイス文のフォーマット ### 【はじめに】＜文章量は2文程度で、良いところを褒めるメッセージを書いてください。＞【課題1に対する評価】＜文章量は1文程度で、学習者の成績の情報から学習者に現状を伝えてください。＞ 【課題2に対する評価】＜文章量は2文程度で、学習者の成績の情報から学習者のエビデンスの選択を評価し、より良いエビデンスの選択を行うためのアドバイスをしてください。＞ 【課題3に対する評価】＜文章量は2文程度で、学習者の成績の情報から学習者の援助内容の選択を評価し、より良い援助内容の選択を行うためのアドバイスをしてください。＞ 【全体を通しての評価】＜文章量は2文程度で、学習者の成績の情報から学習者の現状と今後の伸び代を伝えてください。＞
【文脈情報（看護シナリオに関する情報）】
患者情報 ### ===== ・症例：胃がん ・氏名：武川浩 ・年齢：55歳 ・性別：男 ・経過：年1回、職場の健康診断の他、胃がんの検診のため、胃カメラの検査を受けていた。胃カメラにてポリープが見つかり、精査を進められた。近医を受診し胃カメラにて組織片を採取し、生検の結果、早期胃がんと診断され、内視鏡的切除術による治療目的で入院となった。 =====
【文脈情報（各課題の評価基準となる情報）】
患者問題を判断するエビデンスの重要度の情報 ### ===== ・喫煙歴20年、1日10本喫煙（重要度：3） ・「長年タバコを吸っている」（重要度：3） ・外来受診時肺機能検査 %肺活量80以上、1秒率65（重要度：3） ・医師の所見 肺機能検査にて1秒率において若干の低下があり（重要度：3） ~~~~~（中略）~~~~~ ・「来週から仕事に戻るかな」（重要度：0） ・術後7日目：退院（重要度：0） ・検査の結果：転移の可能性はなく、抗がん剤の使用もしなくても良いでしょう（重要度：0） ・「先生の話聞いて、これですっきりした気持ちで仕事に戻れます。」（重要度：0） ===== ### 患者問題に対する援助内容の重要度の情報 ###===== ・肺合併症のリスク（重要度：3） ・廃用症候群のリスク（重要度：3） ・肺塞栓症のリスク（重要度：3） ・疼痛（重要度：2） ~~~~~（中略）~~~~~ ・胃痛の軽減（重要度：0） ・腸蠕動運動の観察（重要度：0） ・臥床状態による肺炎のリスク（重要度：0） ・痰貯留のリスク（重要度：0） =====

【文脈情報（学習者の成績データ）】
学習者の名前 ### 山下康太 ### 学習者の成績の情報 ### ===== ・課題1. アセスメントに関する課題の評価：5（1~5の段階評価。数値が大きいほど良い。） ・課題2. 患者問題の判断に関する課題の評価：5（1~5の段階評価。数値が大きいほど良い。） ・課題2において選択した患者問題：内視鏡下胃がん摘出術後の合併症のリスク（呼吸器・循環器） ・課題2において患者問題のエビデンスとして選択した項目の一覧：喫煙歴20年、1日10本喫煙、「長年タバコを吸っている」、外来受診時肺機能検査 %肺活量80以上、1秒率65、医師の所見 肺機能検査にて1秒率において若干の低下があり、喫煙にて術後、痰が多くなるなどの悪影響、「煙草は悪いことはわかっているがストレスで止められない」、病室入室後体温36.8℃、病室入室後脈拍80回/分、病室入室後血圧128/78mmHg、病室入室後呼吸17回/分 ・課題2においてエビデンスを選択した理由：患者の喫煙歴が長く、呼吸器疾患を引き起こすリスクが通常に比べて高いと考えたため。 ・課題3. 援助内容に関する課題の評価：5（1~5の段階評価。数値が大きいほど良い。） ・課題3において援助内容として選択した項目の一覧：肺合併症のリスク、廃用症候群のリスク、肺塞栓症のリスク ・課題3において援助内容を選択した理由：術後数日は合併症の発症リスクが高く、経過観察を通してリスク管理を行うことが重要と考えるため。 ・総評：5（課題1, 課題2, 課題3の平均値） =====

表 6 フィードバックの一例

【生成されたフィードバック】
【はじめに】 山下康太さん、あなたの成績についてお話しします。まず、取り組んだ課題については素晴らしい取り組みを見せていただきました。 【課題1に対する評価】 課題1のアセスメントに関する課題については、評価できる成果をお見受けしました。患者の情報に基づいて適切なアセスメントを行うことができましたね。 【課題2に対する評価】 課題2の患者問題の判断に関する課題については、一部改善の余地があります。エビデンスの選択において、より重要な情報にフォーカスすることが求められます。具体的なアドバイスとしては、喫煙歴や肺機能検査結果など、喫煙に関連するエビデンスを重視すると良いでしょう。 【課題3に対する評価】 課題3の援助内容に関する課題については、十分な成果を上げていただきました。適切な援助内容を選択する能力を持っていることが分かりました。 【全体を通しての評価】 山下康太さんの成績は非常に優秀です。課題1では適切なアセスメントを行い、課題3では適切な援助内容を選択することができました。ただし、課題2においてはより重要なエビデンスの選択に注意が必要です。今後も学習を継続し、エビデンスに基づいた看護計画の立案能力を高めていくことが大切です。頑張ってください！

6. まとめ

本研究では、ChatGPT への指示となるプロンプトの設計と生成されるテキストの品質の評価を行った。

評価結果から、成績情報を反映したフィードバックが可能であることが明らかとなった。しかし、表 2 から表 4 に示す課題 1 の評価の精度は、十分な精度が確認できなかった。要因として、プロンプトの総量に対して、課題 1 の成績データに関連する情報の割合が非常に少ないため、ChatGPT が適切に文脈を捉えることができなかったと考えられる。また、表 4 に示す具体的な改善アドバイスについても十分な精度が全ての課題を通じて確認できなかった。要因として、フォーマットで指定したテキスト分量の範囲では、具体的に改善方法を示すことが出来なかったと考えられる。

今後、精度の向上に向け、プロンプトを見直し追加の検証を実施する予定である。

参 考 文 献

- (1) 東サトエ：「看護基礎教育における看護過程の理論モデルと教育プログラムの検討」, 南九州看護研究誌 (2017)
- (2) OpenAI: "Introducing ChatGPT"
<https://openai.com/blog/chatgpt>
- (3) 武田俊之: “生成系 AI の教育および教育研究へのインパクトについて”, 教育システム情報学会研究報告, Vol.38, No.1, pp.51-56 (2023)
- (4) 阿部祐来, 南野早芸, 真嶋由貴恵, 桝田聖子, 川原淳: “大学生のセルフメディケーション実践支援教育に向けた LINE 版おくすり手帳の開発”, 教育システム情報学会研究報告, Vol.38, No.1, pp.17-21 (2023)
- (5) DAIR.AI: "Prompt Engineering Guide"
<https://www.promptingguide.ai/jp>

プログラミングロボットを用いたグループ作品制作における メンバーの役割分担の現状と課題

布施 泉^{*1}, 野口 孝文^{*1}

^{*1} 北海道大学情報基盤センター

Current Status and Issues of Member Role Sharing in Group Work Production using Programming Robots

Izumi FUSE^{*1}, Takafumi NOGUCHI^{*1}

^{*1} Information Initiative Center, Hokkaido University

大学の一般教育として、プログラミングロボットを用いた授業を開講し、様々な知識レベルの学習者が履修をしている。授業では、基礎的な操作を個別学習で学んだ後、グループでロボットの動きを組み合わせた作品を企画し、発表会と相互評価を行う。このグループによる協調作業の中で、著者らはプログラミング能力の育成と深化を期待しているものの、グループによるメンバーの作業分担は必ずしもその期待に沿うものではない場合がある。このような現状と課題について報告する。

キーワード: プログラミングロボット, プログラミング能力, 協調学習

1. はじめに

学習指導要領の改訂で、初等中等教育段階でのプログラミング教育が強化されている⁽¹⁾。この新学習指導要領を経た学生は、2025年度から大学に入学してくる。過渡期を経た後には、全入学生が「プログラミング的思考」を習得して大学に入学することが期待される一方で、学習者の入学時における知識やスキルの幅は大きいことが予想され、その解消は容易ではないことが予期される。

著者らは、2018年度から大学1年生を対象とした少人数の演習授業の中で、プログラミングロボットを用いた授業を行っている⁽²⁾。機械語命令で動作するロボットを操作する中で、プログラミングについて学ぶことを想定して授業を構成している。1年生を対象とする選択の一般教育であるため、文系・理系を問わない学習者がおり、事前の学習者の知識にも幅がある。そのような中で、本授業の授業後の学習者のアンケート調査では、年度を問わず一貫してロボットをプログ

ラミングすることに興味関心が高いことを確認している⁽³⁾。

大学1年生に対するこのような演習系の授業は、初年次教育としても位置付け機能することが望ましいと考える。本授業では、このような視点から2022年度のシラバスでは、到達目標として、以下の3点をあげている。

- ・機械語命令でプログラミングするロボットを用い、ロボットの動きを制御する中で、プログラミングとその背後にあるコンピュータの仕組みを知ることができる。
- ・センサー等の仕組みを用いて複数のロボットの動きを全体企画するなかで、目的に応じてロボットを動かすプログラムを作ることができる。
- ・グループでの学習活動を通し、円滑で効果的な協調活動を行うスキルを獲得する。

まずは、プログラミングロボットを用いてプログラミングについて個別学習により理解を深める。次に、グループで複数のプログラミングロボットを動かす企画

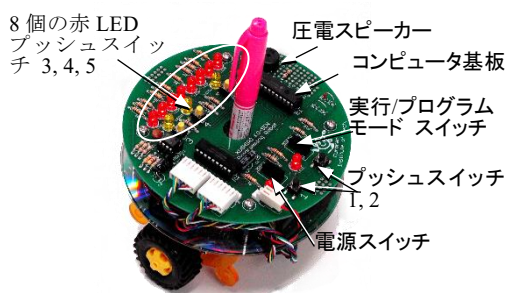


図1 プログラミングロボット

を立てる中で目的に応じたプログラム作成をグループワークとして課し、グループ内でコミュニケーションを深め、協調学習を円滑に進めることを求めている。

著者らの意図は、上記した通りではあるものの、特にセンサーを思いだした 2021 年度から学習者のグループ作品におけるメンバーの役割分担が変化してきた。グループ内の作業分担として、ほぼプログラミングを行わない学習者が発生してきたのである。これはそもそもその到達目標から想定しておらず、どのようにして対処するかを現在検討しているところである。

2 章では本授業の基本的な授業構成、センサーを用いたことによる変化、作品企画における制約条件の変化と作品の状況を述べる。3 章で課題解消に向けた試みを述べ、4 章で現状のまとめを記す。

2. 授業構成と各年度の実践の特徴

2.1 授業構成

本報告における授業は、著者らの所属する大学で実施する大学 1 年生対象のフレッシュマンセミナーという演習で半期 15 回、2 単位のものである。履修者は最大 23 名である。

本授業では、プログラミングロボットを 1 人 1 台用意し、特定のロボットを毎回使用することとしている。学習者は、前半 5 回程度で一通りのロボットの操作を理解する。ロボット単体でもプログラミングは可能であるが、ステップが多い複雑なプログラムは、コンピュータから送信する形の方が容易である。プログラミングロボットの形状を図 1 に示す。命令は 8bit の番地内に格納できる（一部は予約済みで使える番地は 200 程度）。

授業中盤から後半では、3 名から 4 名程度のグルー

表 1 年度によるグループ作業実施の状況の違い

年度	授業形態	PC・協調学習形態
2018	対面 (5 グループ)	専用マシン (1 台/グループ) 作品は 1 台に集約 ロボットは授業時のみ貸出
2019	対面 (6 グループ)	専用マシン (1-2 台/グループ) 人数分のロボットを用いる 協力して描画 ロボットは授業時のみ貸出
2020	対面 & オンライン (6 グループ)	個人 PC (状況により個別貸出) 人数分のロボットを用いる 協力して動作 ロボットは自宅持ち帰り
2021	対面 & オンライン (6 グループ)	個人 PC 人数分のロボットを用いる 協力して動作、センサー利用 ロボットは自宅持ち帰り
2022	対面 (6 グループ、ハイブリット有)	個人 PC 人数分のロボットを用いる 協力して動作、センサー利用 ロボットは自宅持ち帰り可

プに分かれ、グループ発表の作品を企画する。第 13 回か 14 回あたりで発表会ならびに相互評価を行い、第 15 回で授業全体の振り返りを行っている。

2.2 各年度の実践の特徴

2018 年度は履修者が 18 名とやや少なめだが、2019 年度以降は 21-23 名が履修している。2018、2019 年度は対面のみで行い、2020、2021 年度が対面とオンラインを併用、2022 年度は再び対面（体調が悪い学習者が発生した場合には対面と Zoom のハイブリッド）で実施した。

2020 年度と 2021 年度は、グループでの協調学習をオンラインで行うこととし、その間はロボットの持ち帰りを実施した。2022 年度はその流れを受け、各自の進捗状況に応じ、連絡の上持ち帰りも可とした。

表 1 に各年度の実施状況の違いをまとめた。

2018、2019 年度は、PC とロボットの接続には教員側で用意した専用 PC を授業時にグループ毎に貸し出していた（2018 年度は PC5 台 5 グループ、2019 年度は PC9 台 6 グループ）。ロボットも授業時のみ貸し出し、グループとしての作品が完成しない場合には、教

員が付き添うことを条件に時間外学習を行うことを認めた。

2020, 2021 年度は、オンラインを念頭に授業を構成する必要があったため、対面時は個別学習によるロボット操作と個人 PC の環境設定に焦点をあて、残りはウェブ会議システムを用いたオンラインでの授業として構成した。2022 年度になり、コロナも若干落ち着いたため、授業は原則対面に戻した。一部、体調不良もしくは、冬場で JR 不通の際の代替手段として Zoom による出席を認めた。

2021 年度からロボットに光センサーを取り付け、ロボット上にあるプッシュスイッチの押下をセンサーに代替させることができるようになった。ロボットにはプッシュスイッチが 5 つ付いているが、そのうちの 3 つをセンサーに代替させることが可能である。光センサーにより、スイッチ押下時と同じ条件分岐プログラムを実行することができる。2022 年度には、2 つのセンサーを用いてラインレースを行えるようにした。

2.3 グループ作品の制約と制作時の作業分担

2022 年度のグループ作品制作において課した条件は以下の通りである。

- ・ 全てのロボットは 0 番地からプログラムをスタートさせる（最初にあるロボットのセンサーを検知として動作開始）。
- ・ 実行後は、終了まで原則ロボットには触らない
- ・ グループ員全員のロボットを動作させる
- ・ 各ロボットで必ずセンサーを用いる（各ロボットで最低 1 個のセンサーを用いる）※すべてのセンサーを使わなくてもよい
- ・ 音を出す（いずれかのロボットでよい）
- ・ ボールを使う（グループで 1 つ以上）
- ・ これ以外の要素が入っても構わない

グループメンバー分のロボットを動作させるため、3 台もしくは 4 台のロボットで作品を構成する。4 台のロボットがある場合には、1 台を音楽担当にするグループが殆どである。また、残りのロボットは順次、前のロボットの動作に応じて反応する形の作品とすることが多い。2022 年度の発表作品の提示状況を図 2 に

まとめる。

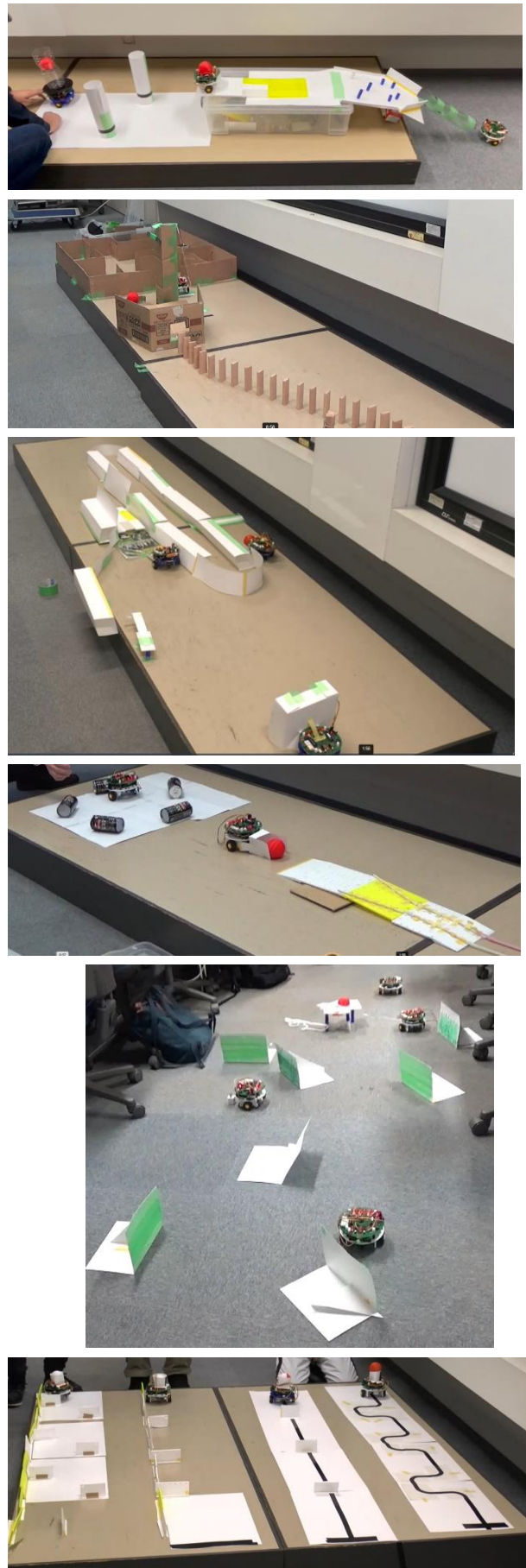


図 2 2022 年度発表会の舞台。6 グループのうち上 5 グループはロボットが直列動作、下は並列に動作。

全体を通して、舞台装置に様々な工夫が凝らされていた。そのため、メンバーの作業分担として、プログラムではなく舞台のみを制作する担当がでてきたグループがある。例えば、図2の上から3番目のグループでは、サーキットを2台のロボットが回るが、基本的に同じ動作をさせ、ほぼ同一のプログラムをロボットに組み込んでいる。協力してそのプログラムを制作していれば授業意図としては問題ないのであるが、結果的に1名が2台のロボットのプログラム担当、1名がサーキットの舞台作り、最後の1台は音楽等別動作という形で分担していた。また、上から2つ目のグループは、最初に迷路があり、最初のロボットは右手の法則で迷路を脱出するプログラムを組み込んでいるが、その後は、前のロボットをセンサー感知してまっすぐ進むだけのプログラムのロボットが繋いでおり、極めて単純なプログラムのみで完了している。しかし、そのような担当は土台に設置した紐をロボットが超えたらその紐が切れ、ボールが飛び出すといったギミックを組み込んでいる。ロボットを動かすプログラムの工夫ではなく、舞台工夫のウエイトが全体として高い傾向が見受けられた。結果として学生の満足度は高いものの、著者らの意図としたプログラミング能力の向上に資する作業分担にはなっていなかったグループがあった。

2.4 グループ作品発表会の成功に向けた改善思考

2022年度のグループ作品では、発表会に成功した事例が少なく、リベンジで及第点の動作をしたグループが殆どであった。これは、一つのロボットの動作のずれが、微妙に次のロボットに引き継がれ、すべてのロボットが100%の出来を出せないことに起因している。例えば、図2の一番下のグループは4台のロボットが別々に動作するものである。発表会ではこのうちの1つが動作しなかったのであるが、その部分だけ再度動作をさせ、グループメンバーは概ねの成功に満足していた。一方でそれ以外のグループでは、どこかでミスが出てしまうと一つの作品として完了しない作りになっていた。

本来は微妙なロボットのずれは、センサーで補正す

るように企画をしていたが、学習者たちの工夫に反映されることは少なかった。

一方、授業で行ったボタンによる条件分岐処理の内容を理解し、一つのボタン（センサー）で、一度目の動きと二度目の動きを変えるようなプログラムを工夫する学習者もいた。何ができるか、どのようにするとできるか等をプログラムの工夫する学習者と舞台装置の改良で問題を解決しようとする学習者に分かれた。

3. 課題解消に向けた試み

2020年度までは、各自のロボットで独立して描画をさせたり、ダンスをさせたりしたため、グループ作品においても各学習者がプログラミング要素を検討することが必要になっていた。しかし、2021年度からセンサーを用いた舞台上での動作を課したところ、前述の通り舞台に凝るグループが発生し、結果としてプログラミングを作業分担として行わない学習者が発生した。このような状況では、誰がどの程度、ロボットにプログラムを入れて実行していたのか、どの程度の理解をしているかの確認も難しくなる。

この状況を改善するために、2023年度はロボットへのプログラム転送状況を保持し、学習者の状況把握をできる独自システムの使用を検討している。

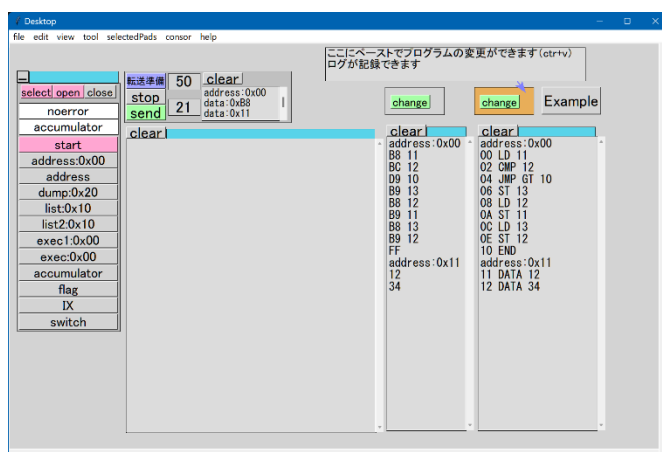


図3 ロボットへのプログラム転送画面とログ蓄積

図3に、学習者の各回の授業におけるロボット入力データを蓄積するための環境を示す。ロボットにプログラムを入力するために個人PC上で用いていたシステムを改修し、ロボットへプログラムを入力し、送信する際（「send」ボタンを押下した際）に、PCの特定ディレクトリ内に送付プログラムを機械語及びアセンブリ言語で蓄積する機能を付与した。日付のディレクトリ内に時間・分・秒が記載されたファイルが格納されていく。

毎回の授業終了時に、PC内にある日付フォルダ内のデータを全てアップロードしてもらうことを想定している。これは教授者にも学習者にもメリットがある。教授者は学習者の細やかなプログラム作成状況の把握が可能になるとともに、ロボットへのプログラム送付変遷を理解することができる。どの段階でどのように困っているかの質問対応にも乗りやすい。また、学習者は、ロボットの動作がうまくいかない場合、時間を確認してプログラムを戻すことができるようになる。これまでは、意識的にプログラムを確保しなければ、なくなってしまう途中プログラムを自動保持できることのメリットは大きいと考えている。

次に学習者の満足度は、制作した作品が100%ではなくとも意図に沿って動作することにより高まると考えられる。前述の通り、2022年度は発表会時の完成度合いは高くなかったものの、リベンジ時間を最大限取することで、どうにか学習者の達成感は確保できていた。しかし、授業時間は限られており、いつも達成できるとは限らない。

もしも1台ずつロボットが順次動作し、それをセンサーで接続していくような(ロボット直列の)動きで、仮に4台続くなれば、各ロボットが9割の成功率であったとしても全体の成功率は0.9の4乗の6割強程度となってしまう。2つのロボットを並列して同時に動かす等、できるだけ各ロボットの出来が次に派生する可能性を減らすような企画が望まれる。

最後に、様々なプログラム例を提示することで、プログラムのできそうなことを増やすことを検討したい。実際にはプログラム例を活用する機会を増やすことが望ましいと考えられるため、グループ活動の前に

ペア活動を挿入し、比較的簡単な課題を共同で解く期間を作ることを検討している。

グループ活動の一環として、舞台装置に凝ることは悪いことではない。しかし、例えば垂直の壁でないと機能しない迷路脱出プログラムの改善方法は、舞台装置としての迷路を完璧に作ることで達成するのではなく、垂直プログラムでない場合にプログラムのどのようになると解決できるかを検討するような取り組みとすることの方が本授業の設計意図としては望ましい。

4. まとめ

本稿では、プログラミングロボットによるグループ作品の企画において、メンバーの作業分担がプログラミング能力育成につながらない場合があることを示した。また、その改善策をいくつか検討している。2023年度の授業実践で、これらの試みがどのように学習者の状況に反映するのかを確認し、実際にプログラミング能力がどのように変化していくかを評価していきたいと考えている。

参 考 文 献

- (1) 文部科学省、平成 29・30・31 年改訂学習指導要領（本文、解説）、
http://www.mext.go.jp/a_menu/shotou/new-cs/1384661.htm(2023 年 6 月 14 日確認)
- (2) 布施泉、野口孝文、梶原秀一、千田和範、稲守栄、“ロボット教材を用いた個別学習を連携した協調学習”，教育システム情報学会研究会報告，vol34，no.2，pp.89-95（2019）
- (3) 布施泉、野口孝文、梶原秀一、千田和範、稲守栄，“プログラミングロボットを活用したハイブリッド型授業における協調学習の実践と評価”，教育システム情報学会研究会報告，vol36，no.2，pp.31-38（2021）

故障予測および作業内容予測機能を持つ カスタマエンジニアの学習支援システムの提案

中村伊吹^{*1}, 左座祐之助^{*1}, 増田武史^{*1}, 広瀬啓雄^{*1}, 三代沢正^{*1}

^{*1} 公立諏訪東京理科大学

Proposal for a customer engineer learning support system
with failure prediction and work content prediction functions.

Ibuki Nakamura^{*1}, Yuunosuke Zouza^{*1}, Takeshi Masuda^{*1}, Hiroo Hirose^{*1}, Tadashi Miyosawa^{*1}

^{*1} Suwa University of Science

保守業務において、問題の原因を特定し対処方法を検討するスキルと、部品交換の適切な時期を予測するスキルが重要となる。本研究ではこれらスキルの獲得を目的とした学習支援システムの提案と、システムの中核となる部品交換予測モデル、作業内容予測モデルの構築を行う。提案システムでは、各予測モデルの予測値を参考に保守業務へ取り組むことで、2つのスキル習得を効率的に行うことを期待している。部品交換予測モデルの構築では、インクヘッド交換の予測を行い、交換日から30日以内に6回中3回のヘッド交換を検出した。作業内容予測モデルの構築では、問題発生時の現象記録から作業カテゴリを予測し、82.3%の正解率であった。今後は各予測モデルの精度向上と提案した学習支援システムを実装する。

キーワード: 学習支援システム, 大判インクジェットプリンタ, 故障予測, 作業内容予測

1. はじめに

A社では大判インクジェットプリンタの製造販売と保守サービスを実施している。大判インクジェットプリンタは、顧客が販売する製品のシールやラベル作成といった業務で利用されている。機器に問題が発生した際には、顧客業務の生産性を低下させる恐れがあり、早急な保守業務の遂行が求められている。A社では保守業務において、保守作業者の減少による業務の遅延と新人育成が問題視されている。

業務の遅延について、A社では問題発生時に顧客からの通報を受け、客先に出向き保守作業を行っている(図1)。通常であれば、問題発生時の通報を受けてから1~3日の間で機器を復旧させることが出来る。しかし、保守作業者が既に他の現場に出向いている等の理由により不在であった場合、機器が復旧するまでに7~10日程経過してしまう。その間、顧客は機器を用いた業務が行えなくなってしまう。そこでA社では業務の遅延を解消するために保守に必要な人員の確保が重要となる。

次に新人育成について、A社での保守業務には必要なスキルが2点ある。1点目は、問題発生時の対処方法を検討するスキルである。対処方法を検討するには、問

題発生時の限られた情報から原因を特定する必要がある。この調査には時間がかかってしまう。ただし熟練作業者は、この作業を自身の経験や勘を基に省略することがある。例えば、機器から異音が聞こえた場合、熟練作業者は過去の似た事例を基に故障原因等を判断し、短時間で対処を行うことが出来る。作業者がこのスキルを習得するには、長期間の実務を通して得られる経験や知識といったノウハウを獲得することが重要である。そのため、十分なノウハウを有していない新人作業者にとって、問題発生時の対策を早急に検討することは困難である。2点目は、事前に問題発生を予測するスキルである。保守業務の際、部品の清掃や設置調整等を行っているが、部品に何らかの異常がみられた場合は対象の部品を交換する。保守業務における部品交換は、A社での負担となるため、部品交換の必要性を見極めることは重要である。熟練作業者は、業務を通して得られたノウハウを基に各部品の状態を見極め、適切なタイミングで部品交換を行うことが可能である。しかし、新人作業者は、部品の状態を把握するノウハウを有しておらず、不適切なタイミングで部品交換を行ってしまう。結果として、本来稼働できたはずの部品が稼働できなくなってしまう、無駄なコストの増加につながる。

いずれも、これら技術の習得には長期間の業務を通して得られるノウハウの獲得が欠かせない。A 社での、ノウハウの獲得には、実務をこなすしかなく、ノウハウの学習に関する有用な学習基盤がない。ゆえに、新人作業者の育成には多くのコストが必要となる。新人作業者が熟練作業者のノウハウを効率的に学ぶための仕組みがあれば、新人作業者の教育にかかる時間を抑え、より短期間で作業者を育成することが可能になる。

しかし、ノウハウの伝承には様々な問題が挙げられる。ノウハウには音や感覚のような言語化の難しい情報が含まれていることや、作業員自身がノウハウに対する自覚が薄いことで、形式知化することが困難であることが多い。ノウハウをテキストや動画で蓄積することが出来たとしても、膨大なデータから必要な情報のみを抽出するためには、多くのコストを要する。また、近年では製造業の現場における技能人材の不足が深刻化しており、熟練作業員が直接ノウハウを伝承しようにも、普段の業務が多忙で技術伝承の時間を確保することが困難になっている。

以上より、本研究で提案する学習支援システムには、熟練作業員が技術伝承にかかる時間を最小限に抑えつつ、新人作業員が効率的にノウハウを学習する機能が求められる。中でも、問題解決のための作業内容を決定するスキルと問題発生を予測するスキルの学習支援が重要である。

そこで、本研究の目的は熟練作業員のノウハウを効率的に獲得するための学習支援システムの提案を行い、システムの中核となる作業内容予測モデルと部品交換予測モデルの構築を行う。

提案する学習支援システムは、作業内容予測機能と故障予測機能からなり、各予測モデルから出力される予測結果を参考に保守業務に取り組ませる。その結果、新人作業員は予測モデルから新たな視点や気づきを得ることができ、短期間で業務に必要なノウハウを身に着けることが期待される。さらに、このシステムを活用することで保守業務の技術支援につながることも期待される。保守業務での技術支援の取り組みとして、故障予測機能では、保守に必要なリソースの確保や予知保全の実施、作業内容予測機能では故障原因の推定等を行う。詳細については3章で述べる。

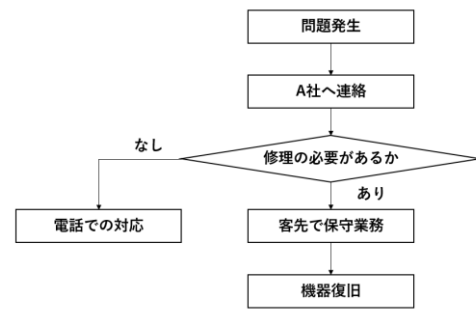


図 1：A 社における保守対応

2. 関連研究

近年の製造業における予知保全では IoT 技術の発展に伴い、工場設備や機器に設置されたセンサーデータに対して機械学習を用いて分析する方法が行われている[2]。データの分析に機械学習を用いる利点として、データの解析や予測において、専門的な知識を必要としないことが挙げられる[3][4]。最近では Automated Machine Learning と呼ばれる、機械学習のモデル構築を自動化するサービスも提供されており、機械学習の活用が容易になってきている[5]。これは、多くの企業にとって、データ分析への敷居を下げ、新たなシステムやサービスの提供に繋がる。

機械学習を用いた故障予測を予知保全へ活用しようとする研究も報告されている。宇宙機の故障予兆検知では、教師なし学習の異常検知を用いて X 線天文衛星の電源系機器の故障予兆検知を試み、一定条件下で故障の予兆を検出することが出来ている[6]。教師あり学習のクラス分類を用いた工場機器の故障予測では、化学プラントに取り付けられたセンサーデータから機械学習を用いてモデル作成、評価実験を行い、発生頻度の高い故障に対して高い予測精度を得ている[7]。これらの先行研究では発生頻度が高い故障への予測や特定条件下での予測には成功しているものの、発生頻度が低い故障の予測精度が低い問題を抱えている。

他にも、本研究で扱うプリンター機器の保守作業を支援するシステムへの取り組みも行われている。[8]のリコーでは全世界の 200 の国と地域に MFP (Multifunction Printer) の販売・サービスを展開しており、日本国内だけでも 400 か所のサービス拠点を有し、数千名の保守担当者が MFP の保守サポートを行っている。リコーでは蓄積されたビッグデータを解析

することで、故障の予兆検知と検知後のアクションを定めたシステム開発を行い、故障予測の現場での活用を実現させている。

教師あり学習を用いた故障予測に関する先行研究では、発生頻度が低い故障の予測精度が低い問題点がある。そこで本研究では発生頻度の高い故障については教師あり学習のクラス分類を用いて予測し、発生頻度の低い故障については教師なし学習の異常検知を用いる。私の知る限り、先行研究ではこの 2 種類の予測モデルを利用した保守作業支援システムに関する研究はない。

次に[8]の先行研究では故障の検出から、現場で必要な作業の確定までを行っている。この先行研究と比較して本研究では故障に関するビッグデータを蓄積できておらず、同様の予測方法では困難である。そこで、本研究では保守作業記録から保守に必要な作業内容を予測することで、故障検出後の保守業務支援を行う。

3. システム設計

1 章で述べたように A 社では、熟練作業者が技術伝承にかかる時間を最小限に抑えつつ、新人作業者が効率的にノウハウを学習する機能が求められる。そこで、本研究で提案するシステムでは、故障予測機能と作業内容予測機能を作成する。そして、予測結果を参考に保守業務に取り組むことで、短期間で業務に必要なノウハウを身に着けることを期待している。加えて、予測結果を保守業務の技術支援につながることも期待している。

故障予測の目標として、故障が発生する 30 日以内に検出することを目標に設定した。これは、故障の検出が早すぎる場合には、本来稼働できたはずの時間を無駄にしてしまう恐れがあるためである。作業内容予測での目標は、予測を活用した保守作業を行うことで、保守業務にかかる時間を 50%削減することが出来れば、熟練作業者のノウハウを効率的に活用できているとして設定した（表 1）。システムフローを図 2 に示す。

表 1：本研究で提案するシステムの機能と目標

機能	目標
故障予測	問題が発生する 30 日以内に検知
作業内容予測	現場での保守業務にかかる時間を 50%削減

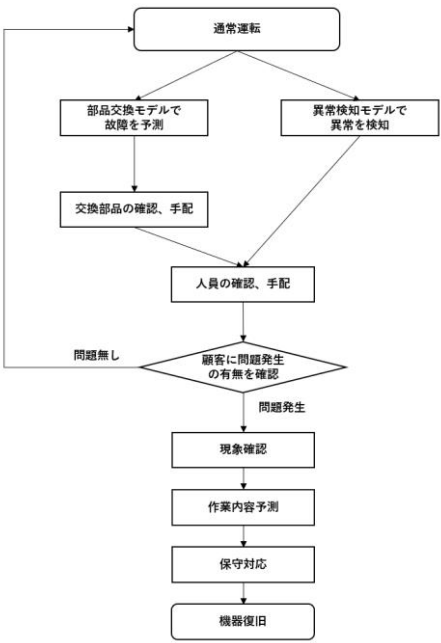


図 2：提案する学習支援システム

・故障予測機能について

本研究で提供する故障予測機能では、異常検知モデルと部品交換予測モデルを作成することで、適切なメンテナンス時期の予測と保守に必要なリソースの確保に活用していく。故障予測機能におけるシステムの構成を図 3 に示す。また、予測モデルの構築方法の詳細については 4 章で述べる。

部品交換予測モデルでは、部品の適切な交換時期の予測と交換部品の調達を行うことが期待される。交換部品の在庫が不足してしまった場合、保守業務の遅延が生じ、顧客の満足度を下げる可能性がある。一方で、交換部品の在庫が過剰にあると、部品の管理に不要なコストをかけてしまう恐れがある。交換部品の需要を正確に把握することが出来れば、計画的に保守に必要なリソースを確保することが可能である。部品交換予測モデルの作成には、交換が必要な部品を特定する必要があり、教師あり学習を用いて予測することが有効であると考えられる。しかし、教師あり学習を用いた関連研究では、発生頻度の高い故障は予測に成功しているが、発生頻度の低い故障においては、データ不足等の要因により予測が行えない問題点がある。そこで、本研究の部品交換予測モデルの構築には、最も発生頻度の高いインクヘッドの交換を対象に行う。

異常検知モデルでは、教師なし学習の異常検知を用いて機器の全般的な異常を検知し、保守作業における人員の手配を行うことが期待される。故障の記録は非

常に少なく、教師あり学習を用いて全ての故障を予測することは困難であるが、教師なし異常検知では、故障記録の少ない異常を検知することが可能である[9]。保守業務に必要な人員の手配には、機器が異常か正常かを判別することが出来ればよい。ため、教師なし学習の異常検知が人員の手配に有効であると考えられる。他にも異常検知モデルでは、部品交換予測モデルで行う予測の信頼性を高める役割を期待している。異常検知モデルと部品交換予測モデルでは、検出可能な問題発生範囲は異なっている（図4）。ただし、2つの予測モデルが検出可能な範囲では、より信頼性の高い予測を行うことが出来ると考えられる。

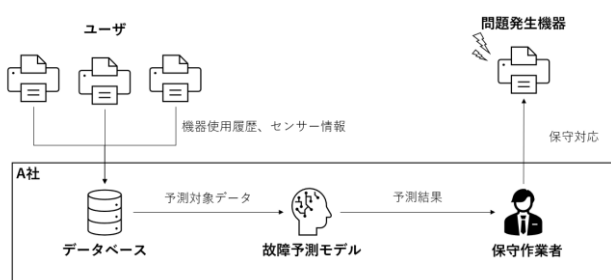


図 3：故障予測機能のシステム構成

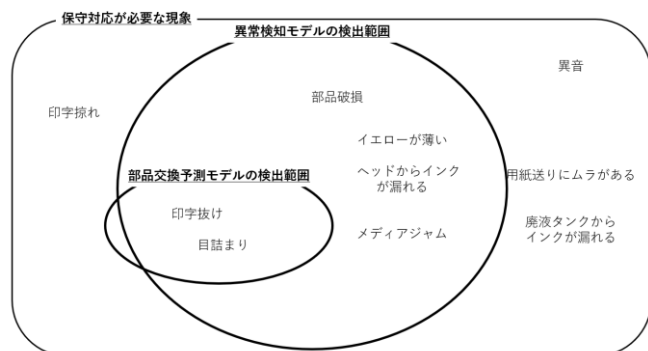


図 4：各予測モデルによる問題発生の検出範囲例

・作業内容予測機能について

作業内容予測機能では、保守作業記録を用いて作業内容の予測を行っていく。保守作業記録には問題発生時の現象、原因、対策が記載されている。本研究では、この保守作業記録が業務におけるノウハウであるとして、問題発生時の現象記録から作業内容を予測するモデルを構築する。このモデルを保守業務の際に使用することで、熟練作業者のノウハウを他の作業員でも活用できるように提供していく（図5）。これにより、技術伝承にかかるコストを削減し、問題発生の解決を効率的に行うことが期待される。



図 5：作業内容予測機能の構成

4. 部品交換予測モデルの構築

4.1 対象データ

本章ではA社のプリンターインクヘッドの交換予測モデルを構築する。インクヘッドとは文字や写真などのデータをもとに、印刷対象にインクを吐出するプリンターのコア部品である。そのため、インクヘッドに問題が発生した際には、印刷に支障をきたすため早急な対応が必要になる。また、インクヘッド交換は保守作業において最も交換回数が多く、最も高額な部品であるため、インクヘッド交換の予測がA社と顧客に与える影響度は非常に大きい。

インクヘッド交換予測モデルの構築には、プリンター機器から不定期に送られてくるプロファイルデータを用いる（表2）。プロファイルデータにはインク使用料といった機器の使用状況が分かる変数が存在する。データの取得期間はおよそ3～4年であり、この期間にヘッド交換が1,2回行われている。また、次の前処理を実施して学習に使用した。

1. サンプル間隔の変更
2. データ数の少ない機体の削除
3. カウンター系変数の初期化
4. 移動平均、階差系列
5. 次元圧縮

表 2：データ概要

	次元数	データ件数	故障件数	機体数
前処理前	125	3748	18	39
前処理後	302	15877	17	14

4.2 部品交換予測モデルの作成

各プリンター機器から送信されてくるプロファイルデータと機械学習を用いてインクヘッド交換の予測を行い、保守業務への有効性を検証する。保守において故障検出が早すぎる場合、機器が本来であれば稼働できたはずの期間を無駄にする恐れがある。そこで本研究では、ヘッド交換が起こる30日以内に交換を予測する

ことを目標に行っていく．予測には教師あり学習のクラス分類を用いる．使用する教師データはヘッド交換日をクラス 1, それ以外の日をクラス 0 として作成した．使用するデータの 70%を訓練データ, 30%をテストデータとする．訓練データには 11 件の交換記録とその間の正常データ, テストデータには 6 件の交換記録とその間の正常データが含まれている．次に実験の手順を示す．

1. 特徴量選択と予測

データの次元数が膨大であるため, 予測に重要な特徴量を選択する．特徴量選択には機械学習の中でもホワイトボックスな手法である Random Forest と Tab Net を使用する．この 2 つと前処理後の全ての変数を用いて交換予測を行い, 各変数の重要度を求める．

2. 特徴量選択後の予測

両手法における重要度の上位 10 個ずつ, 最大で 20 個の特徴量を用いて Gated Recurrent Unit で最終的な交換の予測を行う．

4.3 実験結果

4.3.1 Random Forest による特徴量選択と予測

機体 B, C, E では交換確率がヘッド交換日に近づくにつれて, 徐々に上昇する傾向がみられた(図 6)．テストデータにおいてヘッド交換日に交換確率が 50%を超えたのは 6 回中 3 回で, ヘッド交換日以外に交換確率が 50%を超えたのは 903 回であった(図 7)．変数の重要度については, ヘッド交換日毎に初期化したカウンター系変数の重要度が高い結果となった(表 3)．

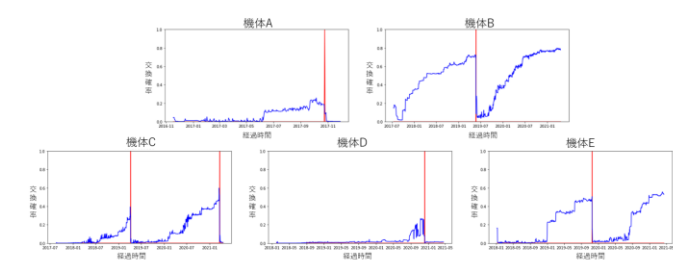


図 6 : Random Forest による予測結果, 縦軸 : 故障確率, 横軸 : 経過時間, 赤線 : ヘッド交換日 (ヘッド交換の行われた日を 1, それ以外の日を 0) 青線 : 予測モデルによって得られた故障確率

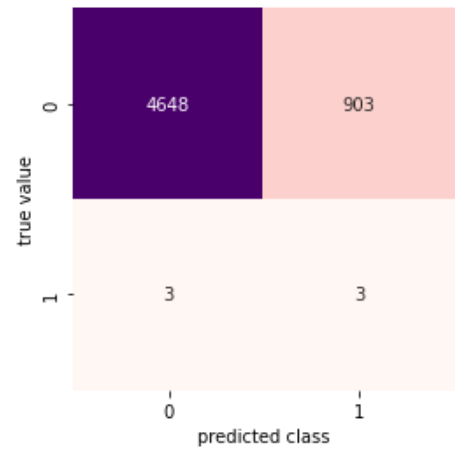


図 7 : Random Forest による予測結果 (混同行列)

表 3 : Random Forest から算出された重要度上位 10

変数	重要度
データ長_cvlog_reset	0.0240
カラー番号5のインク消費量_cvlog_reset	0.0236
カラー番号4のインク消費量_cvlog_reset	0.0233
timestamp_cvlog_reset	0.0228
キャリッジY方向モーター_cvlog_reset	0.0179
ポンプモータAのステップ数_cvlog_reset	0.0171
累積印字タイマ_cvlog_reset	0.0170
原点側廃液のインク量	0.0169
カラー番号2のインク消費量_cvlog_reset	0.0166
カラー番号インク消費平均_cvlog_reset	0.0160

4.3.2 Tab Net による特徴量選択と予測

ヘッド交換日に近づくにつれて交換確率が徐々に上昇する傾向はみられなかった(図 8)．テストデータにおいてヘッド交換日に交換確率が 50%を超えたのは 6 回中 2 回で, ヘッド交換日以外に交換確率が 50%を超えたのは 361 回であった(図 9)．変数の重要度では, Random Forest と同様に, ヘッド交換日毎に初期化した変数に加えて, 一週間ごとの移動平均をとった変数の重要度が高かった(表 4)．

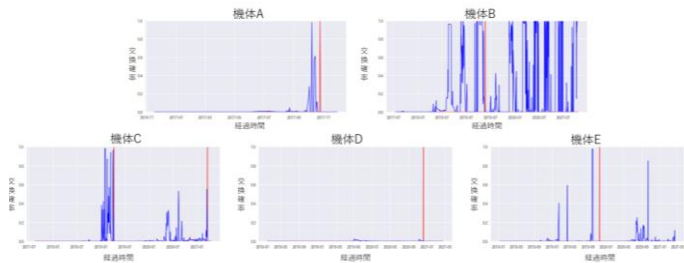


図 8 : Tab Net による予測結果, 縦軸 : 故障確率, 横軸 : 経過時間, 赤線 : ヘッド交換日 (ヘッド交換の行われた日を 1, それ以外の日を 0) 青線 : 予測モデルによって得られた故障確率



図 9：Tab Net による予測結果（混同行列）

表 4：Tab Net から算出された重要度上位 10

変数	重要度
カラー番号インク消費合計_cvlog_reset	0.2815
ヘッド1インク消費合計_mean	0.1721
ポンプモータ寿命：累積_cvlog_reset	0.1536
ポンプモータ寿命：累積_mean	0.1072
ヘッド1~4インク消費合計_mean	0.1049
ヘッド2インク消費合計_mean	0.0693
カラー番号5の充填されているインクの種類	0.0564
Isomap_neighbors15_1	0.0414
原点側廃液のインク量_mean	0.0043
カラー番号5のインク消費量_mean	0.0005

4.3.3 Gated Recurrent Unit による予測

予測に用いる変数は Random Forest と Tab Net から得られた重要度の上位 10 個ずつを用いた。機体 C,D,E では交換確率が交換日に近づくにつれて、徐々に上昇した(図 10)。テストデータにおいてヘッド交換日に交換確率が 50%を超えたのは 6 回中 5 回で、ヘッド交換日以外に交換確率が 50%を超えたのは 1232 回であった（図 11）。

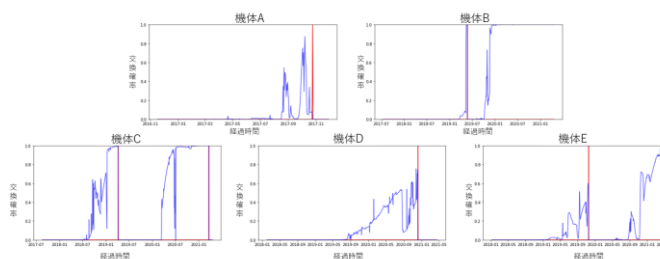


図 10：Gated Recurrent Unit による予測結果,縦軸：故障確率,横軸：経過時間,赤線：ヘッド交換日（ヘッド交換の行われた日を 1,それ以外の日を 0）青線：予測モデルによって得られた故障確率

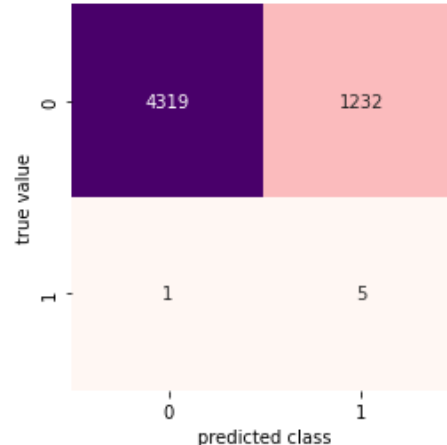


図 11：GRU による予測結果（混同行列）

4.4 考察

今回行った予測では,交換日までに交換確率が 50%を超えたものの,その後確率が下降してしまうことがある.そのため,一度でも交換確率が 50%を超えた時点からヘッド交換までの日数を,交換検出期間とする.表 5 に各機械学習手法によるインクヘッド交換の検出期間を示す.

表 5：各機械学習手法による検出期間の比較

	機体A 1回目	機体B 1回目	機体C 1回目	機体C 2回目	機体D 1回目	機体E 1回目
Random Forest	×	403	×	3	×	1
TabNet	215	×	310	372	×	180
Gated Recurrent Unit	23	11	201	354	135	5

重要度の高い特徴量について、ヘッド交換日毎にカウントをリセットした変数が Random Forest と Tab Net において重要度が高かった.理由として次の2点が考えられる.1 点目は,今回使用したデータに欠損値が多かったため,補間を行いやすいカウンター系変数の重要度が高いと判断されたこと.2 点目は,カウンター系変数をヘッド交換日ごとに初期化したことで,インク消費量の値やモータ寿命値の組み合わせがインクヘッドの経年劣化を表せたと考えられる.

本予測の目標はヘッド交換日の 30 日以内に故障を検知することである.Gated Recurrent Unit ではヘッド交換日の 30 日以内に 6 回中 3 回のヘッド交換を検出した.また,A 社で働く社員に今回の予測に対する評価をいただいた結果,Gated Recurrent Unit を用いた故障予測はメンテナンスに有効であると判断された.また,検出期間があまりに長すぎるものは,異常検知モデルから算出された異常度との比較を行うことで予測

に対する信頼性の向上を図っていく。

4.5 システム改善の方策

本研究で使用したデータは予測モデル構築のために取得したものではない。そのため今後は次の方策を進めることで、システム実装に向けた予測モデルの精度向上とデータ取得基盤の改善に取り組む。

1. データ取得間隔の改善

使用したデータは不定期に取得しているため、機体によっては一か月に一度しか取得していないこともある。データの取得間隔をより短くなれば、短期的に変動する変数を扱えるようになるため、直近の予測を行うことが期待される。

2. 特徴量選択手法の改善

本研究での特徴量選択は、Random Forest と Tab Net に依存している。そのため、A 社職員に故障原因のヒアリングをし、予測に有効な変数を調査することで精度を向上することが期待できる。

3. 汎用性の評価

A 社で働く社員の方に今回の予測に対する評価をいただいた結果、本研究で使用したデータでは、メンテナンスに有効であると判断された。本実験で構築した予測モデルを活用して保守業務を行うには、新たに取得したデータを用いて予測精度を調べ、汎用性の評価を行う。

5. 作業内容予測モデルの構築

5.1 対象データ

作業内容の予測には保守作業記録を用いて行う。データ概要は表 6 に示す。保守作業記録には問題発生時の現象、原因、対策、作業内容カテゴリが記録されている。本実験で使用する保守作業記録は 72 件のみで、予測対象の作業カテゴリの件数にはばらつきがある。そのため、本実験での予測対象は作業件数が、上位 3 つの作業カテゴリに限定する。予測を行う作業カテゴリ毎のデータ件数は表 7 に示す。

表 6：保守作業記録の概要

現象	原因	対策	作業カテゴリ
印刷時の擦れ	ヘッドの不具合	ヘッド交換	部品交換
メディアにインクが垂れる	ヘッド周りの汚れ	クリーニング	清掃
...	...	...	...

表 7：作業カテゴリ毎の発生件数

部品交換	設置	清掃
37	12	7

本実験の目的は、技術伝承にかかるコストの削減を行う予測モデルを構築することである。予測モデルの構築には自然言語処理モデルの BERT を用いる。学習には説明変数として問題発生時の現象、目的変数を作業カテゴリとして教師あり学習を行う。また、使用するデータの 70% を訓練データ、30% をテストデータとしてモデルの構築を行う。

5.2 作業内容予測モデルの作成結果

表 8 にテストデータにおける予測結果を示す。作業カテゴリの予測を行った結果、テストデータにおける予測の正解率は 82.3% であった。作業カテゴリにおいて清掃の必要があるものを部品交換と予測してしまう結果となった。

表 8：テストデータによる予測結果

		予測結果		
		部品交換	設置	清掃
正解データ	部品交換	11	0	0
	設置	1	3	0
	清掃	2	0	0

5.3 考察

今回の実験では問題発生時の現象から作業カテゴリを予測し、テストデータにおける正解率が 82.3% であった。また、清掃における予測では 1 件も正解することが出来なかった。清掃の予測が出来なかった原因として、学習に使用したデータに含まれる清掃の件数が最も少なかったことが考えられる。他にも、現象の記録は保守作業者が自然言語で記述したものであるため、記載内容には個人差がある。解決策として、記録の記載をチェックシートにすることで、作業者の個人差によらず、均一の作業記録を取得することが期待できる。

本実験の目的は作業カテゴリの予測によって技術伝承にかかるコストの削減を行うことである。今回は作業カテゴリの予測を行ったが、作業カテゴリの予測だけでは、どの部品に対して作業を行うのかが不明であるため、保守業務の支援につながる可能性は低いことが考えられる。今後、保守業務の支援を行うには作業カテゴリを予測するだけではなく故障の原因や作業箇所

の特定を行う必要がある。

5.4 システム改善の方策

本研究で使用したデータは予測モデル構築のために取得したものではない。そのため今後は次の方策を進めることで、システム実装に向けた予測モデルの精度向上とデータ取得基盤の改善に取り組む。

1. データの新規取得, データ拡張

今後も新規の保守作業記録を取得し、清掃の予測に十分なデータの確保を行っていく。データ拡張に関しては、保守作業記録に記載されている単語をもとに、学習用データを拡張することで、予測精度の向上が期待される。また、作業者が過去に行った保守業務の内容をヒアリングすることで、保守作業記録にはないノウハウを取得することが期待される。

2. チェックシートの作成

保守作業記録の記載内容に個人差があることが原因で、予測精度が低下したことが考えられる。そのため、現時点での現象からチェックシートを作成することで記載内容の均一化を行っていく。また、作業記録をカテゴリ変数にすることで、データ数が少なくても学習可能な手法の適用が期待される。

3. 作業箇所, 故障原因の予測

本実験での予測では作業カテゴリの予測を行ったが、保守業務を行うには、故障原因の特定と対策が重要である。そのため、本実験で実施した作業カテゴリの予測だけでは、業務を行うためには不十分である。今後は、問題の原因や対策を予測していくことで、保守業務に活用可能な予測モデルを実装する。

6. まとめ

本研究では、熟練作業者のノウハウを効率的に獲得するための学習支援システムの提案と、システムの中核となる部品交換予測モデルと作業内容予測モデルの構築を行った。

部品交換予測モデルの作成では、Random Forest, Tab Net から重要度の高い特徴量を選択し、Gated Recurrent Unit で予測することで、交換日の30日以内に6回中3回の交換を検出した。また、Gated Recurrent Unit による予測をA社の職員に評価していただき、メンテナンスに有効であると判断された。精度を向上す

るためにはデータ取得間隔の変更やA社職員のヒアリングを通じて特徴量を精査する必要がある。

作業内容予測モデルの作成では、問題発生時の現象から作業カテゴリの予測を行い、正解率は82.3%であった。また、清掃に関する予測は1件も正解することが出来なかった。原因として、清掃に関する記録が7件しかなく、予測対象である作業カテゴリの中で最も少なかったことや、作業記録の記載に個人差が大きいことが挙げられる。予測精度を向上する方法として、保守作業記録をチェックシートに変更することで、記載の個人差を少なくすることが期待される。また、本実験では作業カテゴリの予測を行ったが、保守業務を支援するためには作業箇所, 故障原因の予測を行う必要がある。

本研究で使用したデータは、予測モデルの構築を目的に収集したものではない。そのため、今後はシステム実装に向けて、今回の結果を基にデータ取得基盤や予測精度の改善に取り組んでいく。また、異常検知モデルの作成と図2に示した保守システムの実装と評価を行う。

参 考 文 献

- (1) 経済産業省, “製造基盤白書 (ものづくり白書)”, 2018
- (2) 今村 誠, “予知保全のための機械学習”, システム/制御/情報, 65 巻, 4 号, p.119-125, 2021
- (3) Yann LeCun, Yoshua Bengio, and Geoffrey Hinton: Deep learning. Nature 521.7553 2015: 436-444
- (4) Rockart, J. F. and Morton, M.S.S.: “Computers and the Learning Process in Higher Education”, McGraw Hill, New York (1975)
- (5) 松尾豊, 人工知能は人間を超えるか: ディープラーニングの先に あるもの: KADOKAWA, 2015.
- (6) Google: Cloud AutoML, <https://cloud.google.com/automl>
- (7) 梅津里香, 杉江卓哉, 長瀬雅之, et al, “機械学習を用いた宇宙機の故障の予兆検知, 宇宙科学情報解析論文誌”, JAXA-RR-18-008, 8 号, p.11 - 20, 2019
- (8) 切通 恵介, 泉谷 知範, “機械学習を用いた工場機器の故障予測”, DEIM forum 2017
- (9) 奥 武憲, 水野 哲志, 林 俊輔, 荒井 仁, et al. MFP のビッグデータを活用した故障予測”, Ricoh technical report, 43 号, p.56-62, 2018
- (9) 曾我部東馬: Python による異常検知, オーム社, 2021

Scratch ブロックの多様な組み合わせを考慮した プログラム自動採点方法の開発

畠中 明哉^{*1}, 広瀬 啓雄^{*1}, 尾崎 剛^{*1}

^{*1} 公立諏訪東京理科大学大学院

Development of an Automated Grading Method for Scratch Programs Considering Diverse Combinations of Blocks

Akiya HATANAKA^{*1}, Hiroo HIROSE^{*1}, Takeshi OZAKI^{*1}

^{*1} Suwa University of Science

This study developed an automated grading system for Scratch programs. The system extracts the required information from the solution programs, compares them with the model answers to determine correctness, and displays the results. By validating the constrained problems with multiple expected answers, it was confirmed that accurate grading is possible.

キーワード: プログラミング教育, ビジュアルプログラミング, Scratch, 初等教育

1. はじめに

日本では, IT 人材不足の解消を目的に, 2020 年度から小学校でのプログラミング教育が導入された. 文部科学省が公開している「小学校プログラミング教育の手引(第三版)」¹では, 「プログラミング的思考」の育成や各教科等での学びをより確実なものにすることなどがねらいとして挙げられている. 小学校のプログラミング教育は, 多様な教科等に取り入れることが可能であり, そのためには児童がプログラミングに必要な知識・スキルを身につけるための学習活動を行う必要がある. しかし, その学習活動を各教科等に組み込んでしまうとプログラミング教育の本来の目的が達成されない可能性がある. そのため, 教員が授業時間をやりくりしてプログラミングを学ぶ時間を確保しているが, 十分な時間を確保することができない可能性がある. そこで, 児童が短時間で効率的な学習を可能にし, どの程度プログラミング的思考やプログラミングに関する知識・スキルを習得したか確認するための方法を明らかにする必要がある.

Scratch などのビジュアル型プログラミング言語を用いたプログラムを対象に, 学習者が作成したプログラムに対して自動で分析や評価を行う研究は既に行われている. 太田ら²は, 小学校におけるプログラミング教育での教師や児童の利用を想定しており, 児童が自由に作成したプログラムに対して, 分析や評価を行うための基準を新たに提案し, その基準に基づいてプログラムの評価を行う学習支援システムを開発した. このシステムはプログラムの評価を行うものであり, 特定の課題に対して児童が解いた問題の自動採点を行うシステムではない.

著者らは, プログラムの効率的な学習用教材として児童が数分程度取り組める Scratch ドリルの開発を目指している. これまで, プログラミングを学ぶ課題に対して, 作成されたプログラムを正しく採点する正誤判定方法の提案と検証を行ってきた³. しかし, 正しく判定を行うことができなかったブロックの組み合わせが存在したため採点方法の再検討が必要であった.

これらをうけて本研究は, 学習者が問題に対して作成したプログラムを正しく判定できる正誤判定システ

ムの構築を目的に行った。

2. プログラム自動採点システム

2.1 Scratch

現在、小学校では Scratch などのビジュアル型プログラミング言語を使った学習指導が展開されている。Scratch は、米国のマサチューセッツ工科大学メディアラボと Scratch 財団が共同開発したビジュアル型プログラミング言語である。ブロックタイプのコマンドを組み合わせることでプログラムを作成することができる。Scratch 開発環境の画面は大きく 3 つに分けられている。画面左はブロックパレット、画面中央にスクリプトエリア、画面右がステージ及びスプライト情報が表示されている。ブロックパレットには、動き・見た目・音・イベント・制御・調べる・変数・ブロック定数の 8 種類に分けられたブロックがある。これらのブロックをスクリプトエリアにドラッグ&ドロップし、組み合わせることでプログラムを作成する。プログラムの実行結果がステージ及びスプライトに表示される。

さらに、Scratch にはプログラムを拡張できるものとして拡張機能がある。2019 年に公開された Scratch3.0 では、micro:bit や翻訳、音声合成などの外部機器や API を扱う拡張機能が追加された。また、拡張機能は Scratch 公式が用意しているものの他に、拡張機能を自作することができ、Stretch3⁴ のような第三者が作成した拡張機能を搭載した Scratch システムが Github 上で公開されている。本研究では、Scratch3.0 をつかったシステムを構築し、以下、Scratch3.0 を Scratch と表記する。

2.2 自動採点の流れ

自動採点は Scratch の拡張機能を用いて実現する。図 3 はプログラムの自動採点の流れである。学習者が拡張機能を選択すると、図 4 のように Scratch のブロックパレットに問題ごとに作成された正誤判定ブロックが追加される。問題に対する解答を作成した後、対応する正誤判定ブロックを先頭に配置することで正誤判定が行われる。正解不正解の判定結果がステージ上

に表示される。

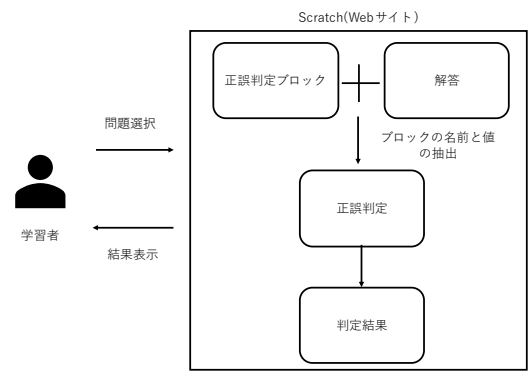


図 1 自動採点の流れ

2.3 採点機能の構築

Scratch のソースコードは、Github 上で公開されている。公開されているソースコードのうち、scratch-gui と scratch-vm をダウンロードし、拡張機能の構築に関係のあるファイルを編集することで、拡張機能を自作することができる。本研究では、公式の Japanese Scratch-Wiki⁵ の拡張機能の作り方を参考に、新たな拡張機能として、正誤判定を行うオリジナルブロックを作成していく。また、Scratch の開発言語は JavaScript であるため、オリジナルブロックの動作を記述する際も JavaScript を使用する。

表 1 ブロックの種類と形状

種類	形状
真偽ブロック	
スタックブロック	
ハットブロック	
C 型ブロック	
値ブロック	
キャップブロック	

Scratch で利用可能なブロックを形状別でみた場合、表 1 に示す 6 種類がある。これらのブロックのうち、ハットブロックは、その下にブロックが追加される度

に処理が行われる。この特徴は、学習者が作成したプログラムの正誤判定を行うのに適しているため、本研究では正誤判定ブロックをハットブロックで作成した。また、値ブロックの中には、選択するとそのブロックに代入されている値をステージ上に表示することができるブロックがある。このことから、判定結果の表示は値ブロックで作成した(図2)。



図2 正誤判定ブロック・判定結果ブロック

2.4 正誤判定

正誤判定は、解答プログラムに使われたブロック名、値、順序が模範解答と一致するかどうかで行う。

2.4.1 Scratch のブロック

学習者が作成したプログラムを採点する上で、Scratch のブロックから必要な情報を読み取る必要がある。Scratch にはブロックパレット、スクリプトエリアにある全ブロックの情報をもつ BlockUtility オブジェクトというプロパティがあり、ステージ上のスプライトの座標、向き、大きさやスプライトの動きのメソッドなどの情報をもっている。スクリプトエリアに作成したプログラムの情報も BlockUtility に存在する。BlockUtility からスクリプトエリアに作成したプログラムを解析すると、どのブロックが使用された読み取ることができる。また、ブロックにはブロックパレットに置かれているブロックの他にも、値を入力・変更することができるブロックなどは、隠しブロックと呼ばれる値を格納しているブロックが存在する。

Scratch のブロックには、id・opcode・inputs・

fields・next・topLevel・parent・shadow・x. y というプロパティが格納されている。この中で正誤判定に必要なプロパティは id・opcode・inputs・fields・next プロパティである。id プロパティはスクリプトエリアに置かれているブロックに振り分けられている独自の id である。opcode プロパティはブロック名を表しており、例えば、「○歩動かす」ブロックの opcode は、「motion_movesteps」である。inputs プロパティは隠しブロックが存在するブロックや、「ずっと～する」や「もし～なら」などの繰り返し処理・条件分岐処理のブロックがもっており、隠しブロックの値や繰り返し・条件分岐処理のブロックの条件やその中に置かれているブロックを知りたい場合は inputs を調べると分かる。fields プロパティは隠しブロックの値が格納されており、入力された値を調べる際に有効である。next プロパティは、次に置かれているブロックの id をもっており、ブロックの並び順は next をたどることで分かる。

2.4.2 解答プログラムの読み取り

解答プログラムからブロック名、値、順序を自動的に読み取る機能を構築する。学習者が作成したプログラムからブロック名と入力された値を抜き出し、ブロック名を key、入力された値を value としたプロパティを、オブジェクトに格納する。ブロックの名前は opcode を、入力された値は隠しブロックの fields を使う。



図3 「○歩動かす」ブロック

図3の「○歩動かす」ブロックを例に説明する。「○歩動かす」ブロックの opcode は「motion_movesteps」で、入力されている値は 10 である。これを {"motion_movesteps":10} としてオブジェクトに格納する。



図4 多様なブロックの組み合わせ

しかし、Scratch では多様なブロックの組み合わせで同じ動きをすることが可能である。図4は、図3のブロックと同じ動きをするブロックの組み合わせの例で、(1)ブロックを複数置いた組み合わせ、(2)「〇回繰り返す」ブロックを使用した組み合わせ、(3)演算ブロックを使用した組み合わせである。これらも図3の場合と同様に、`{“motion_movesteps”:10}`となるようにした。



図5 プログラム例1

図5は、「動く・待つ・動く」のような複数の動作をするプログラムの例である。このようなプログラムでは、ブロックの順序を取得する必要がある。ブロックの順序は、ブロックの `next` プロパティを辿ることで取得することができる。図5のプログラムは、間に違う種類のブロックを挟んで、同じ種類のブロックが配置されている。これらのブロックの順序を明確にするために、opcode に番号をつけて格納する(図5右)。

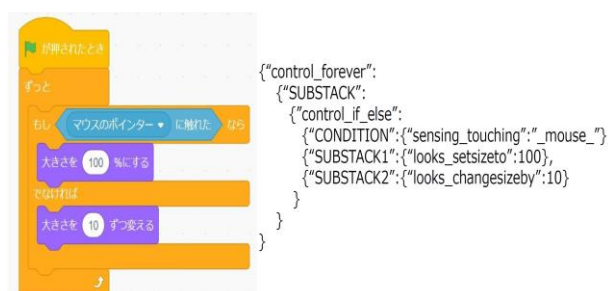


図6 プログラム例2

図6は、繰り返し処理と条件分岐処理のブロックが組み込まれたプログラムである。繰り返しや条件分岐などの制御ブロックの中などに置かれているブロックは、`inputs` プロパティを使って読み取ることができる。

「`control_forever`」は、「ずっと～する」ブロックを指している。このブロックのように中に置けるブロック群が1つの場合は、`inputs.SUBSTACK` プロパティを調べることでプログラム群の読み取りが可能となる。

「`control_if_else`」は「もし～なら、でなければ～」ブロックを指す。このブロックは条件ブロック、条件が真の場合のブロック群、および条件が偽の場合のブロック群を置くことができる。これらは、「ずっと～する」ブロックと同様に `inputs` プロパティを使って調べる。`inputs.CONDITION` プロパティから条件ブロック、`inputs.SUBSTACK1` プロパティから真の場合のブロック群、`inputs.SUBSTACK2` プロパティから偽の場合のブロック群を読み取ることが可能である。

以上のような処理を行うことで、解答プログラムから必要な情報だけを抜き出し、オブジェクトに格納することができる。

2.4.3 正誤判定

正誤判定のために模範解答を用意する。問題に対する模範解答プログラムを上記の方法でオブジェクトに格納したものを模範解答とする。

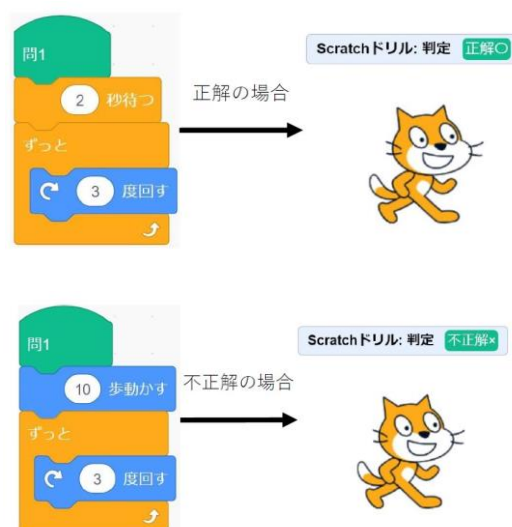


図7 判定結果の表示

図7のように正誤判定ブロックに解答プログラムを

組み合わせることで、解答が模範解答と一致するかどうかを判定する。解答が模範解答と一致していれば「正解○」、そうでなければ「不正解×」と表示する。

3. 検証

検証用の問題とそれに対する正誤判定ブロックを 7 つ作成し(表 2)、正しく判定を行うことができるかどうかを検証した。

表 2 検証用の問題

	問題文
問 1	1 秒待った後に、時計回りに 3 度回し続けなさい。
問 2	もしスプライトがマウスポインターに触れたなら 3 歩動かしなさい。そうでなければ 2 歩動かしなさい。そして、もし端に着いたら跳ね返るようにしなさい。これらをずっと繰り返しなさい。
問 3	スプライトの大きさを 100%にしなさい。そして、スプライトの大きさを 10 ずつ変え、もしマウスポインターに触れたなら大きさを 100%にする動きをずっと繰り返しなさい。
問 4	スプライトの大きさを 100%にしなさい。そして、1 秒待ちスプライトの大きさを 5 ずつ変え、1 秒待ち・スプライトの大きさを 50 ずつ変える動きをずっと繰り返しなさい。
問 5	1 秒待ち、スプライトを 0 度に向けなさい。次に 1 秒待ち、スプライトを 90 度に向けなさい。
問 6	スプライトを時計回りに 15 度回し 1 秒待ち、時計回りに 90 度回し 1 秒待つ動きを 3 回繰り返しなさい。
問 7	スプライトがマウスポインターに触れている時はずっと時計回りに 5 度回しなさい。

検証用の問題は、1 つのスプライトかつ 1 つのスク립トでプログラムを作成できる問題に限定し、「プログラミング能力検定 過去問題集 ビジュアル言語版 レベル 1」⁶の 6 月分の問題を参考に、プログラムの基

本制御構造である、順次・分岐・反復を含む問題となるように作成した。

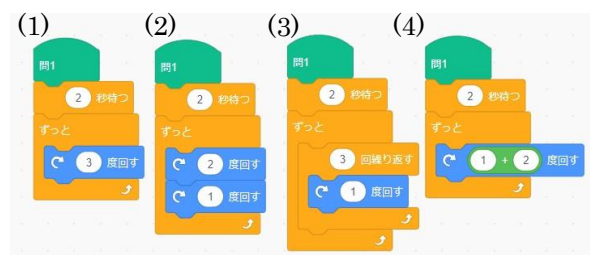


図 8 正解の想定解答

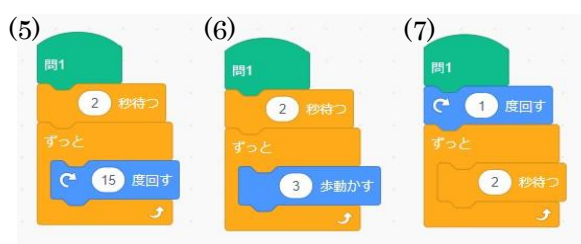


図 9 不正解の想定解答

検証用の問題に対して、正解の解答には、図 8 のような (1)「模範解答」、(2)「ブロックを複数置いた解答」、(3)「『○回繰り返す』ブロックを置いた解答」、(4)「演算ブロックを用いた解答」、不正解の解答には、図 9 のような、(5)「数値が異なるブロックを置いた解答」、(6)「ブロックの種類が異なる解答」、(7)「ブロックの順番が異なる解答」の想定解答を用意した。(2)から(7)の想定解答においては、ブロックの数値やブロックの組み合わせのパターンを複数用意した。これらの想定解答を用いてシステムの検証を行った。その結果、すべての問題に対し、用意したすべての想定解答で正しく判定を行うことができた。

4. 考察

本研究では、1 つのスプライトかつ 1 つのスク립トで作成することができる 7 つの問題で検証を行った。その問題に対し、正解・不正解の想定解答を複数用意し検証を行った結果すべて正しく判定を行うことができた。このことから、構築した正誤判定システムは自動採点が可能であるといえ、この条件に合致するプログラムを学ぶ際に利用可能であると考えられる。

Scratch のプログラムには、検証問題のようなプロ

グラム以外にも、変数を利用するプログラムや、1 つのスプライトに対して 2 つ以上のスクリプトで作成するプログラム、複数のスプライトに対して 1 つまたは複数のスクリプトで作成するプログラムもある。今回構築したシステムはこれらのプログラムに対応しておらず、このようなプログラムでの検証は行っていないが、これらに対応した正誤判定システムの実現は十分に可能であると考え、開発を進めている。

5. まとめ

本研究は, Scratch で作成されたプログラムの自動採点システムの開発を行った。解答プログラムから、正誤判定に必要なブロック名、値、順序を抜き出す機能を構築した。その抜き出した情報が、模範解答と一致しているかどうかを判定し、結果を表示するシステムを開発した。

本システムで正しく採点を行うことが可能か検証するために、1 つのスプライトかつ 1 つのスクリプトでプログラムを作成することができるプログラムに限定した問題を 7 つ作成した。その問題に対し、正解・不正解の想定解答を複数用意し検証を行った結果すべて正しく判定を行うことができた。

しかし、今回の研究では、限定された条件下の問題であり、変数を利用するものや 2 つ以上のスクリプトを使うものなどで正しく判定を行うことができるか検証できていない。

児童が短時間で効率的に学びことができる教材として Scratch ドリルシステムの開発し、活用してもらうためには、小学生が学ぶべきプログラミング要素を整理し、各要素を段階的に学習できる問題を作成する必要がある。問題の作成にあたっては、実際に小学生に取り組んでもらい、難易度・学習効果・問題の受容性について検証を行う。最終的には Scratch ドリルシステムを小学生にを使って学習を行ってもらい、プログラミング的思考やプログラミングに関する知識・スキル習得度合いを確認するための方法として有効であるか検証する予定である。

謝辞

本研究は、JSPS 科研費 22K02893, 23K02715 の助成を受けたものです。

参 考 文 献

- (1) 文部科学省：「小学校プログラミング教育の手引(第三版)」, (2020)
- (2) 太田 剛, 加藤 浩, 森本 容介：“コンピューショナル・シンキング概念に基づくプログラム自動評価機能を持つ Scratch 用学習支援システムの開発”, 教育システム情報学会誌, Vol. 35, No. 2, pp204-214, (2018)
- (3) 畠中 明哉, 尾崎 剛, 広瀬 啓雄：“Scratch ドリルシステム開発に向けた Scratch プログラム自動採点方法の検討”, 教育システム情報学会 2022 年度学生研究発表会 A06, (2023)
- (4) Stretch3,
<https://github.com/stretch3/stretch3.github.io>
(2023 年 6 月 13 日確認)
- (5) Scratch3.0 の拡張機能を作ってみよう-Japanese Scratch-Wiki,
https://ja.scratch-wiki.info/wiki/Scratch_3.0の拡張機能を作ってみよう (2023 年 6 月 13 日確認)
- (6) プログラミング能力検定協会 著, 『2022 年度版 プロ検® 過去問題集 ビジュアル言語版 レベル 1』, 株式会社スプリックス, 東京都, (2022)

野菜摂取における講義とベジチェックがもたらす大学生の健康 行動変容への影響

～スタディエンゲージメントと自己効力感の調査から～

柳谷 泰志^{*1}, 阿部 祐来^{*2}, 真嶋 由貴恵^{*2}, 榎田 聖子^{*2}

^{*1} 大阪府立大学 現代システム科学域

^{*2} 大阪公立大学大学院 情報学研究科

Influence of lectures on vegetable intake and veggie-check on health behavior changes of university students

～From the Survey of Study Engagement and Self-Efficacy～

Taishi YANAGITANI^{*1}, Yuki ABE^{*2}, Yukie MAJIMA^{*2}, Seiko MASUDA^{*2}

^{*1} College of Sustainable System Sciences, Osaka Prefecture University

^{*2} Graduate School of Informatics, Osaka Metropolitan University

The health literacy of young people in Japan is low, and they tend to eat less vegetables, especially in their meals. There are also reports that many young people have low self-efficacy. Therefore, we gave a lecture on the necessity of vegetable intake to university students and measured the effect of the lecture with a Veggie-Check every week, and investigated whether there was an increase in awareness of vegetable intake and behavioral change in terms of study engagement and self-efficacy. As a result, the improvement of the Veggie check value and the change of awareness and behavior were observed.

キーワード: 野菜摂取, ベジチェック, 健康行動変容, スタディエンゲージメント, 自己効力感, 大学生

1. はじめに

野菜にはビタミンやミネラルなどの栄養成分が多く含まれるとともに, 食物繊維やポリフェノールなど, がんや糖尿病をはじめとする生活習慣病を予防する多様な成分も含まれる⁽¹⁾. このように, 野菜は健康の保持増進のために欠かせない食品であり, 厚生労働省は, 「21 世紀における国民健康づくり運動 (健康日本 21 (第 2 次))」において, 野菜の平均摂取量として 350g/日を目標値としている⁽²⁾. しかし, 令和元年国民健康・栄養調査結果⁽³⁾では, 成人の野菜の平均摂取量は 280.5g/日に留まり, 特に大学生が該当する 20～29 歳では男女ともにどの年代よりも少なく, 222.6g/日と報告されている.

一方, 少子高齢化が進み生産年齢人口の減少が課題となっている中, 企業内ではワークエンゲージメントの向上に関心が集まっている. 令和元年版労働経済分析によると, 5 年前と比較すると, 日本において「エンゲイジメント」を含む記事数は約 10 倍となっており, 急速に関心が高まっている傾向が確認できると報告されている⁽⁴⁾.

ワークエンゲージメントとは, 仕事に関連するポジティブで充実した心理状態であり, 活力, 熱意, 没頭によって特徴づけられるものと定義されている⁽⁵⁾. 一方, 学生におけるワークエンゲージメントとは, 学びへの意欲や関心, 取り組み方を包括した考え方であり, これを高めることが学生の学習成果の向上や成長に繋

がと考える⁽⁶⁾。本研究では、学生におけるワークエンゲージメントをスタディエンゲージメントと定義する。

エンゲージメントを高める要素の一つとして、メンタル面に対する「情緒的サポート」が重要だと言われている⁽⁷⁾。一方で近年の調査では、諸外国と比較して日本の若者の自分自身への満足度や肯定感が低いことが指摘されており、また、自身の可能性を認知する「自己効力感」も低く、本来の能力を発揮できないといったことが起こっている⁽⁸⁾。そこで、スタディエンゲージメントや自己効力感を高めることが必要である。これらを高めるためには、健康的な日常生活が基本となり、まずは食生活を整えることが有効であると考え。身体の状態は精神の状態と密接に関係しており、食生活に関する多くの先行研究では、食生活の改善がメンタルヘルスの改善や予防に繋がる可能性があることを結論付けている⁽⁹⁾。

そこで本研究では、大学生を対象に適切な野菜摂取を促し、これにより自己効力感やスタディエンゲージメントが向上するかを調査する。これをきっかけに学生が適切な食習慣を身に着けるような健康行動変容につながれば、将来の自己効力感やワークエンゲージメントの向上につながるだけでなく、健康増進による健康寿命の延伸にも寄与すると考える。

2. 研究手法

野菜摂取に関する知識の獲得と食習慣の改善・定着を目指す。具体的には、野菜に関するセミナーを行った後、週に1回、自身の野菜摂取量を測定できる「ベジチェック」を活用し、その値の変化を確認する。

2.1 野菜に関するセミナー

K社の管理栄養士により60分程度のセミナーを実施する。今回は主に、野菜に関する知識や1日の目標野菜摂取量(350g)と効率的な野菜摂取法について教授する。

2.2 ベジチェックによる評価

ベジチェック(カゴメ株式会社)とは、手のひらをセンサーに約30秒あてるだけで、簡単に推定野菜摂取量を見える化できる機器である。推定野菜摂取量は0~12の値で表示され、適切な摂取量とされる350g以上⁽²⁾摂取している場合は7.0以上の値が表示される。



図1 ベジチェック測定方法

3. 実験方法

3.1 対象者

本学の現代システム科学域知識情報システム学類の専門科目「ヘルスケアサービス」(2単位・選択)の受講生に対し、研究の趣旨、目的などを文書で説明し、参加に同意の得られた者4名を対象者とする。

3.2 実験期間

実験期間は、2022年11月25日から2023年1月24日までの約2ヶ月間である。

3.3 評価指標

適切な野菜摂取を促し、これにより自己効力感やスタディエンゲージメントの向上が見られるかを以下の2項目で評価する。

I：実験期間中のベジチェック値の推移

II：実験前後における自己効力感・スタディエンゲージメント値

3.4 倫理的配慮

本研究は、大阪公立大学情報学研究科における倫理委員会の承認を得て実施する。対象者のデータは学籍番号で管理する。対象者については科目担当教員には知らせず、対象者に不利益がないように配慮する。

3.5 実験手順

対象者への実験参加の同意を得たのち、図2に示すように①事前アンケート、②食習慣改善ワーク、③事後アンケートの流れで行う。

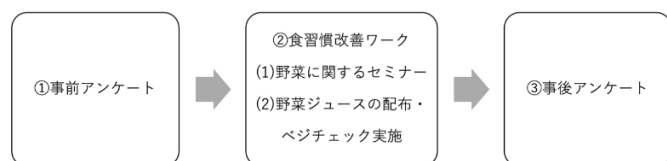


図2 実験手順

①事前アンケート

実験前の対象者の野菜摂取状況、自己効力感(3項目)、ワークエンゲージメント(14項目)に関するアンケート調査を実施する。事前アンケート項目を表1に示す。

②食習慣改善ワーク

野菜摂取を促すワークを(1)(2)の手順で行う。

(1)野菜に関するセミナー

対象者(受講生)は授業内で野菜に関するセミナーを受講する。

(2)ベジチェック実施・野菜ジュースの配布

毎回の授業後に「ベジチェック」の値によって野菜の摂取状況を確認する。Google フォーム上に測定値を記録する。

③事後アンケート

実験後の対象者の野菜摂取状況、自己効力感(3項目)、ワークエンゲージメント(14項目)に関するアンケート調査を実施する。事後アンケート項目を表2に示す。

表1 事前アンケート項目

	設 問	設 問 数
1	研究への同意	1
2	学籍番号	1
3	性別	1
4	年齢	1
5	同居人数	1
6	プレゼンティズム・アブセンティズム	2
7	野菜を 350g 摂る意識	1
8	主観的健康感	1
9	健康管理に関する自己効力感	1
10	ヘルスリテラシー	5
11	スタディエンゲージメント	14
12	自己効力感	4
	合 計	33

表2 事後アンケート項目

	設 問	設 問 数
1	学籍番号	1
2	性別	1
3	年齢	1
4	同居人数	1
5	プレゼンティズム・アブセンティズム	2
6	ヘルスリテラシー	5
7	野菜を 350g 摂る意識	1
8	主観的健康感	1
9	健康管理に関する自己効力感	1
10	スタディエンゲージメント	14
11	自己効力感	4
12	健康管理への影響	1
	合 計	33

4. 結果

3.3 で示した評価指標に基づいて結果を示す。

4.1 ベジチェック値

各週におけるベジチェック値の結果を以下の表 3 に示す。対象者 4 名のすべてにおいて、実験開始時の測定値より終了時の測定値の方が高かった。

表 3 ベジチェック値

	0 週目	1 週目	2 週目	3 週目	4 週目
A	5.2				
B	4.0	4.3	5.3	4.2	5.7
C	7.7	8.1	8.3	8.4	9.0
D	4.3		4.4	4.6	
平均	5.3	6.2	6	5.7	7.4

	5 週目	6 週目	7 週目	8 週目	平均
A				7.0	6.1
B	6.4	5.8	6.8	7.0	5.5
C	8.6	9.0	9.3	9.3	8.6
D	5.3	5.9	6.1	6.0	5.2
平均	6.8	6.9	7.4	7.3	6.5

4.2 自己効力感とスタディエンゲージメント

事前・事後アンケートで調査した自己効力感とスタディエンゲージメントの結果を以下の表 4,5 に示す。

事後アンケートまで完了した対象者は 4 名であった。

自己効力感において、4 個の質問に対して 7 段階評価での回答結果（28 点満点）を以下に示す。A,B,D において自己効力感の低下、C において向上していた。

表 4 自己効力感（28 点満点）

	事前 アンケート	事後 アンケート	前後比較
A	24	22	-2
B	16	15	-1
C	19	22	+3
D	21	19	-2

スタディエンゲージメントにおいて、14 個の質問に対して 7 段階評価での回答結果（98 点満点）を以下に示す。A,B,C においてスタディエンゲージメントの向上、D において低下していた。

特に C は自己効力感、スタディエンゲージメントの両方が上昇、逆に D は両方が低下していた。

表 5 スタディエンゲージメント（98 点満点）

	事前 アンケート	事後 アンケート	前後比較
A	57	61	+4
B	35	40	+5
C	59	64	+5
D	72	68	-4

5. 考察

実験前後で被験者のベジチェック値が上昇していることから、野菜摂取に関するセミナーによって野菜摂取の意識改革につながったのではないかと考える。また、最終計測時において適切な野菜摂取（ベジチェック値 7.0 以上）と判定された A,B,C においてスタディエンゲージメントの上昇、適切な野菜摂取までは至らなかったと判定された D はスタディエンゲージメントの減少が見られたことから、適切な野菜摂取とスタディエンゲージメントには関係性があることが示唆された。これは、食習慣の改善によるメンタルヘルスの向上が、スタディエンゲージメントの向上に寄与した可能性があると考えられる。一方、自己効力感、スタディエンゲージメントの両方が上昇した C 以外では、適切な野菜摂取を行ったからといって、自己効力感の向上につながるとは言えない結果になった。

しかし、自己効力感に影響を及ぼす要因は他にも考えられるため、適切な野菜摂取が自己効力感の向上につながらないと断定することはできないと考える。自己効力感に影響を及ぼす他の要因の例として、実際に自分で実施してうまくいった「成功体験」や、自分と同じ境遇のモデルを見て自分でもうまくできそうだと感じる「代理体験」があり、自己効力感の向上に寄与すると言われている⁽¹⁰⁾。

また、本実験中において、事後アンケートを実施した時期は期末テスト期間であったため、その時期の体調が影響した可能性も考えられる。

6. まとめ

本研究では、大学生を対象に適切な野菜摂取の知識を教授し健康行動変容につなげることで、自己効力感やスタディエンゲージメントが向上するかを調査した。

その結果、野菜摂取に関するセミナーとベジチェックの値でフィードバックすることによって野菜摂取の知識が健康行動変容につながっていくことが示唆された。

また、適切な野菜摂取とスタディエンゲージメントの向上には関係性があることが示唆された。一方で、適切な野菜摂取が自己効力感の向上につながるとは言えない結果になった。

しかし、自己効力感に影響を及ぼす要因は他にも考えられるため、適切な野菜摂取が自己効力感の向上につながるような仕組みを考えることが必要である。

事前アンケートから事後アンケートまで通して実験に協力してくれた対象者は4名と少なかったことから、今後対象者数を増やして正確なデータ分析を行う予定である。

謝辞

本研究において、野菜摂取に関するセミナーの実施、野菜ジュースの提供等、お力添えをいただいたK社の管理栄養士に感謝する。

参 考 文 献

- (1) 池上幸江,梅垣敬三,篠塚和正,江頭祐嘉合:”野菜と野菜成分の疾病予防及び生理機能への関与”,栄養学雑誌, 61(5):275-288(2003)
- (2) 厚生労働省:”健康日本21”
https://www.mhlw.go.jp/www1/topics/kenko21_11/b1.html#A15 (2023年6月6日確認)
- (3) 厚生労働省:”令和元年国民健康・栄養調査”
<https://www.mhlw.go.jp/content/001066903.pdf>
(2023年6月12日確認)
- (4) 厚生労働省:”令和元年版 労働経済の分析 第3節「働

きがい」をもって働ける環境の実現に向けた課題について”,

<https://www.mhlw.go.jp/wp/hakusyo/roudou/19/dl/19-1-2-3.pdf> (2023年6月13日確認)

- (5) 厚生労働省:”第2-(3)-1図 ワーク・エンゲージメントの概念について”,
<https://www.mhlw.go.jp/stf/wp/hakusyo/roudou/19/backdata/2-3-01.html>(2023年6月5日確認)
- (6) Jun Tayama, Wilmar Schaufeli, Akihito Shimazu, Masanori Tanaka, Akari Takahama:”Validation of a Japanese Version of the Work Engagement Scale for Students - Tayama - 2019 - Japanese Psychological Research-WileyOnlineLibrary”,
<https://doi.org/10.1111/jpr.12229>(2023年6月5日確認)
- (7) 向江 亮:”ワーク・エンゲージメント向上の実践的取組に向けた知見の整理と今後の展望”,https://doi.org/10.3222/jaiop.32.1_55(2023年6月5日確認)
- (8) 内閣府:”特集 今を生きる若者の意識～国際比較からみえてくるもの～”
<https://www8.cao.go.jp/youth/whitepaper/h26gaiyou/tokushu.html> (2023年6月5日確認)
- (9) 片山 友子, 水野(松本) 由子, 稲田 紘:”大学生の生活習慣とメンタルヘルスの関連性”,
<https://doi.org/10.7143/jhep.41.283>(2023年6月5日確認)
- (10)eヘルスネット:”セルフ・エフィカシーを高めるポイント”:<https://www.e-healthnet.mhlw.go.jp/information/exercise/s-07-002.html> (2023年6月6日確認)

AI が作成したソーシャルワーク実習記録の考察文章の評価

～学生が作成した文章との比較分析～

坂本毅啓^{*1}, 佐藤貴之^{*1}, 中原大介^{*2}

^{*1} 北九州市立大学, ^{*2} 福山平成大学

Evaluation of AI-generated social work practice record discussion sentences

～Comparative analysis with student-generated writing～

Takeharu Sakamoto^{*1}, Takayuki Sato^{*1}, Daisuke Nakahara^{*2}

^{*1} The University of Kitakyushu, ^{*2} Fukuyama Heisei University

In social work practice in social work education, students work on creating a practice journal (practice record). We prepared two texts about this practice journal, one written by AI and the other written by students during their practice. We then obtained data evaluating those two sentences from two perspectives: the way they were written as practical training records and the content of the learning. The results showed that there was no difference between the AI-generated and student-generated discussion sentences, and that they reached the same level.

キーワード: ソーシャルワーク教育, 実習記録, 文章生成 AI, 人工知能

1. はじめに

2023 年に入ってから, OpenAI 社が提供を始めた対話型の文章作成サービスである ChatGPT は, 社会的に大きな波紋を上げたという指摘はすでにニュース等で多くなされてきている. その影響は教育の現場においても同様である. 現在の教育現場において, AI (artificial intelligence: 人工知能) とどのように向かい合うのか, 活用していくのか, それとも排除していくのか, そもそも教育の目的と実践はこれからどのようにあるべきなのかが, 今まさに問われているとも言える.

筆者らが専門とするソーシャルワーク教育に目を向けると, 2020 年からの新型コロナウイルス感染症 (COVID-19) の拡大を受けた非常事態宣言の影響 (コロナ禍) により, 遠隔授業の実施, ICT (Information and Communications Technology) を活用したオンラインによる代替実習の実施など, それまでは ICT の活用に対して極めて消極的であったそれまでの状況⁽¹⁾から一変したと

言える. コロナ禍における福祉人材の養成に対してガイドライン案が取りまとめられているが, それを踏まえるとソーシャルワーク教育において想定される ICT 活用とは, 実習巡回指導, 実習記録 (実習日誌), その他関連する実習事務である⁽²⁾. 実習での経験とそこからの学びを記録する実習記録に焦点化すると, 「手書きで作成すべき」とされてきた実習記録について, コロナ禍の影響により ICT の活用が一気に進められた (進めざるを得なかった) と言える⁽³⁾.

AI による自動文章作成機能が高度化している一方で, ソーシャルワーク教育における実習記録も手書きから ICT を活用したものへと変化してきている点を踏まえると, 実習記録の作成において学生が AI を活用することは十分に起こり得ると考えられる. しかし, ソーシャルワーク教育において AI を活用した教育の在り方に関する研究は管見の限り見当たらない. 実習記録において AI の活用・規制に関するガイドラインも作成されているとは言えない. 実習記録は単なる経

験の記録（日記）ではなく、経験を振り返り、社会福祉学の専門性を踏まえて考察し、そこからの学びを言語化することを目的としている。自ら考え、言語化する過程で AI を活用することは、ソーシャルワーク教育における専門性の獲得とその発展に負の影響を与えてしまうのでは無いかと、これまでの AI に対する懸念から推察をする。以上から、実習記録の中の実習日誌の考察文章に焦点化することにする。

2. 研究の目的

以上を踏まえて、本稿では実習記録と AI 活用のあり方に関する研究の最初の手掛かりをつかむことを目指して、以下の 2 点の問いについて明らかにすることを目的とする。

- ① AI は実習記録の考察文章を作成することができるのか？
- ② AI が作成した文章と学生が作成した文章はどのような違いがあるのか？

3. 研究方法

3.1 ソーシャルワーク教育における実習記録

ソーシャルワーク教育における実習記録とは、個人票から終了後の「まとめの考察」まで、実習に関連して作成される文書の総称である。その内、日々の経験を記録し、その日の実習目標に基づいた評価・考察、あるいはソーシャルワーカー（社会福祉士）の専門性に基づく評価・考察をするのが実習日誌である。このような様式は標準的なテキストや教員研修で 사용되는テキスト等でも紹介されている様式であり、ICT を活用した実習記録システムでも同様の様式となっている。この実習日誌の大半を占めるのが考察欄である。

以下では、この考察欄の文章（以下、考察文）を AI に作成させ、記録の文章として評価した場合、教育の目的に合致した内容のレベルに到達するのか、学生が作成した文章とどの相違点があるのかを明らかにしていくことにする。以下はその手順である。

- ① 実習を行った学生 1 名に対して、初期段階の実習記録の文章の提供と、同じシチュエーションで ChatGPT に文章を作成させることを依頼する。
- ② 実習日誌の作成における評価項目を作成する。

- ③ 「①」で提供された 2 つの文章を「②」で作成した評価項目で評価（点数化）する Web アンケートを作成する。どちらの文章が AI 作成によるかは伏せる。
- ④ 福祉領域で実戦経験のあるスペシャリストを対象に、「③」で作成した Web アンケートに基づいた実験への協力を依頼する。
- ⑤ 「④」で収集したデータを、関連する 2 群の平均値を比較することで、AI 作成の文章の完成度及び学生作成と AI 作成の文章の相違点を明らかにする。

3.2 評価するための文章の作成

学生に実習 3 日目の記録を提供してもらい、同じシチュエーションでの実習記録の考察文も、同じ学生が ChatGPT を使って考察文を作成してもらった。

実習 3 日目としたのは、実習記録にあまり書き慣れていない段階の記録文章と比較したかったためである。あまり書き慣れていない段階で AI に頼った場合、AI の方が上手に考察文章を作成してしまうと、学生は AI に依存してしまう可能性が高くなる。その結果、学生は実習記録を書く中で、記録を書くスキルの獲得と、記録を書くことを通して専門性を獲得するという 2 点の学習ができなくなるのでは無いかと考えた。以下は、学生が提供してくれた学生が作成した文章と AI が作成した文章である。

【学生が作成した文章】

C利用者についてだが、午後の活動中（よさこい）に手をひっぱられることがあった。その時点でC利用者とはまだ数回しか会話したことがなく、何を伝えようとしてくれていたのか理解することができなかった。後から職員の方に話を聞くと、「床に落ちていた食べのこしが気になっていたのだと思う」とのことだった。C利用者はその場所まで私を連れて行ってくれていたが、やはりすぐには分からなかったため、会話を交わすだけでなく観察し、利用者が何を伝えようとしているのかという背景を考え続けることが重要だと感じた。

続いて、今日の午後の活動を始める際に受けた説明について、新しい発見があったので、記載しておく。午後の活動でパットゴルフを実施したが、その際に「午後

の活動は目的や大事にしている点がいくつかあり、利用者にケガをさせないこと、気分転換となるように盛り上げること、そして職員の近くに配置する人は視野が狭かったり動くことが難しかったりする人であること。」と説明を受けた。配置について考えたことはなく、利用者 1 人 1 人の様子によって内容や実施方法を変更、鑑みることも大事なのだと学んだ。明日の実習についても利用者 1 人 1 人を見て、そこから得られる情報を持ち、どのように活用していくかを考えていきたいと考えている。

【AI が作成した文章】

本日の実習では、利用者の発信に注目し、その意味を考えることが目標であった。この目標に基づいて、清掃中に食堂にいるメンバーや職員との交流を通じて、利用者が発信する意味について考える機会を得た。また、午前の個人ワークの見守りを通じて、利用者の発信にも注意を払いながら観察を行った。

午後の活動では、よさこいとパットゴルフが行われたが、利用者の中には動きが止まってしまう人や他のことに気を取られる人がいた。こうした状況において、どのように声掛けをするかを考えることが求められた。前者については、動きを促すことができたが、後者については、発信していた意図を汲み取ることができなかった。このように、利用者が発信する内容や意図を正確に理解し、適切に対応するためには、コミュニケーション能力が必要であることを再認識した。

今回の実習を通じて、利用者が発信する意味に注目し、それに対する適切な対応を考えることが重要であることを学ぶことができた。明日以降も、この視点を念頭に置きながら実習に取り組みたいと思う。

参考として学生が AI に文章を作成させた際にどのように問いかけたのか、原文をそのまま記載する。

清掃をしながら食堂にいるメンバーと交流した。待ち時間に職員の方に色々質問させていただいた。午前の活動が始まり、昨日同様個人ワークの見守りを行った。本日の目標を念頭に置きつつ利用者を観察することに努めた。

午後の活動はよさこいとパットゴルフだった。よさこい中にも動きが止まってしまう利用者や他のことが気にな

る利用者がいた。それぞれどのように声掛けをすればいいのか自分の中で考え、前者は動きを促すことができたが後者が発信していた意図については汲み取ることができなかった。明日以降の課題である。

という内容でソーシャルワーカーの実習記録の考察分を作成して下さい。

3.3 評価指標の作成

実習記録の書き方、書くべき内容についてテキスト、基本書等を参考にして評価指標を作成した⁽⁴⁾⁽⁵⁾⁽⁶⁾⁽⁷⁾⁽⁸⁾⁽⁹⁾。書き方に関する 10 項目、考察文章の内容に関する項目 9 項目、5 件法（できていない(1)、あまりできていない(2)、どちらも言えない(3)、少しできている(4)、できている(5)）とした。以下は評価項目である。

【文章の書き方】

1. 「常体(文末に「だ、である」を使う)」で文章が書かれている。
2. いつ、どこで、だれが、なにを、どうしたのか、具体的に客観的事実が書かれている。
3. 要点が簡潔に書かれている。
4. 分かりやすい表現で書かれている。
5. 適切な専門用語や略語が使われている。
6. 1文の長さを 35 字以内になっている。
7. 必要時、利用者の言葉がそのまま書かれている。
8. 他から得た情報は情報源も書かれている。
9. 誤字・脱字がない。
10. 不適切な表現がない。

【文章の内容】

1. プライバシーへ配慮して、名前のアルファベット表記など匿名化されて書かれている。
2. その日の実習目標に照らして、その場面を詳細に記述し、考察がされている。
3. 体験した内容を踏まえて、その日の実習目標がどの程度達成されたのか、達成されなかったのか書かれている。
4. 上記の達成された・されていないと判断した理由について書かれている。
5. 判断の根拠となる客観的な事実が書かれている。
6. 要因を分析して具体的に書かれている。

7. 知識や技術など、専門性に基づいて考察をすることが出来ている。
8. 自立支援、自己選択・自己決定等の視点に基づいて書かれている。
9. 次の課題を明らかにすることが出来ている。

3.4 調査方法データ収集

2023年3月18日に、縁故法に基づき、介護教員講習会に参加している受講者（介護、医療、福祉領域の実践者、教員等）44名に対して Web アンケート方式で協力を依頼した。その結果、38名から回答があった。分析は Jamovi (ver.2.3.24) 及び「Statcel4」⁽¹⁰⁾を用いた。

質問項目は回答者の基本属性（年齢、現在の所属、実習日誌の指導経験の有無）、AI の考察文に対する評価指標（19項目）、学生の考察文に対する評価指標（同19項目）とした。なお、AI が作成した文章が含まれているかは伏せた状態で考察文の評価を依頼した。

3.5 倫理的配慮

調査にあたっては協力依頼状を講習時間外に配信し、調査の主旨、アンケート回答の有無・内容等が受講評価とは関係が無いこと、及び倫理的配慮について説明をした上で各人の自由意志に基づいて調査への協力を求めた。調査の実施及びデータの取り扱いについては、北九州市立大学における「人を対象とする研究に関するガイドライン」に則った。

4. 研究結果

4.1 回答者の属性

回答者の属性とその概要については、表1と表2の通りである。無効票は無く、回答率は86.4%であった。

回答者の年齢は平均 42.8 ± 7.87 歳であり、介護教員講習会の受講者に調査への協力をお願いした影響からベテランと言える年齢層が中心となったと言える。現在の所属については、教育機関（大学・短大・専門学校等）が71.1%（27人）、福祉・介護の臨床現場（医療機関等を含む）が28.9%（11人）であった。実習日誌の指導経験の有無については、指導を行ったことがあるが68.4%（26人）、行ったことがないが31.6%（12人）であり、半数以上は実際に実習日誌の指導を行っ

た経験があった。これらを踏まえると、今回の回答者は、年齢的にも教育指導経験的にもある程度の経験をもったベテラン領域にある人が多かったと言える。

表 1 回答者の年齢

年齢	
N	38
欠損値	0
平均値	42.8
中央値	42.5
標準偏差	7.92
最小値	26
最大値	58

表 2 回答者の所属と指導経験

現在の所属	度数	全体%	累積%
教育機関（大学・短大・専門学校等）	27	71.1%	71.1%
福祉・介護の臨床現場（医療機関等含む）	11	28.9%	100.0%
実習日誌の指導経験	度数	全体%	累積%
実習日誌の指導を行ったことがある。	26	68.4%	68.4%
実習日誌の指導を行ったことがない。	12	31.6%	100.0%

4.2 合計点の比較

表 3 学生作成と AI 作成別に見た平均比較

項目	群	N	平均値	中央値	標準偏差	標準誤差
書き方合計点	学生作成	38	38.3	40	7.86	1.27
	AI作成	38	40.2	40	7.02	1.14
内容の合計点	学生作成	38	34.2	35	7.39	1.2
	AI作成	38	35.7	37	8.11	1.32
合計点	学生作成	38	72.5	74	14.39	2.33
	AI作成	38	75.9	78	14.38	2.33

表 4 ウィルコクソン符号付順位和検定の結果

学生作成	AI作成	検定方法	統計量	p	平均値の差	差の標準誤差
書き方合計点	書き方合計点	ウィルコクソンのW	210	0.054	-3	1.34
内容の合計点	内容の合計点	ウィルコクソンのW	181	0.123	-2.5	1.1
合計点	合計点	ウィルコクソンのW	195	0.031	-5	2.24

注: H₀: μ_{測定値1} - 測定値2 ≠ 0

学生作成と AI 作成の考察文章について、書き方の評価項目群と内容の評価項目群ごとの合計点と、さらにそれを合計した合計点（総計）について、学生作成と AI 作成の対応のありの平均値の比較を、ノンパラメトリックのウィルコクソン符号付順位和検定（Wilcoxon signed-ranks test）を用いた。その結果（表3、表4）、書き方及び内容に関する項目群ごとの合計

点については、いずれの差も有意では無かった。

書き方の合計点と内容の合計点を合計した合計点（総計）については、学生作成と AI 作成の平均点の差は有意だった（ $W=195$, $p=0.031$, 両側検定）。総合的に見ると、AI 作成の方が学生作成よりも実習日誌の考察文章として評価が高かった。

4.3 評価項目ごとの比較

各評価項目の記述統計については表 5 の通りである。各項目について正規性の検定（Shapiro-Wilk test）を行ったところ、全ての項目の P 値が 0.05 を下回ったため、全ての評価項目において正規分布とは異なっていると判定した。これを踏まえ、以下の各評価項目の平均値の比較（t 検定）ではノンパラメトリックを用いることとした。

表 5 全項目の記述統計

									シャピロ=ウィルク検定		
		N	欠損値	平均値	中央値	標準偏差	最小値	最大値	W	p	
学生作成	書き方	「常体（文末に「だ、である」を使う）」で文章が書かれている。	38	0	4.32	5	1.165	1	5	0.649	<.001
		いつ、どこで、だれが、なにを、どうしたのか、具体的に客観的事実が書かれている。	38	0	4.11	4	1.06	1	5	0.791	<.001
		要点が簡潔に書かれている。	38	0	3.74	4	1.131	1	5	0.856	<.001
		分かりやすい表現で書かれている。	38	0	4.03	4	1.078	1	5	0.774	<.001
		適切な専門用語や略語が使われている。	38	0	3.55	4	0.921	1	5	0.885	<.001
		1文の長さを35字以内になっている。	38	0	3.16	3	1.48	1	5	0.875	<.001
		必要時、利用者の言葉がそのまま書かれている。	38	0	3	3	1.47	1	5	0.879	<.001
		他から得た情報は情報源も書かれている。	38	0	4.16	4	0.973	2	5	0.793	<.001
		誤字・脱字がない。	38	0	4.13	4	1.018	1	5	0.765	<.001
		不適切な表現がない。	38	0	4.13	4	1.119	1	5	0.75	<.001
	文章の内容	プライバシーへ配慮して、名前のアルファベット表記など匿名化されて書かれている。	38	0	4.37	5	1.125	1	5	0.618	<.001
		その日の実習目標に照らして、その場面を詳細に記述し、考察がされている。	38	0	3.74	4	1.032	1	5	0.884	<.001
		体験した内容を踏まえて、その日の実習目標がどの程度達成されたのか、達成されなかったのかが書かれている。	38	0	3.53	4	1.156	1	5	0.887	0.001
		上記の達成された・されていないと判断した理由について書かれている。	38	0	3.45	3.5	1.132	1	5	0.893	0.002
		判断の根拠となる客観的な事実が書かれている。	38	0	3.87	4	1.07	1	5	0.79	<.001
		要因を分析して具体的に書かれている。	38	0	3.95	4	0.985	1	5	0.844	<.001
		知識や技術など、専門性に基づいて考察をすることが出来ている。	38	0	3.71	4	0.956	1	5	0.832	<.001
		自立支援、自己選択・自己決定等の視点に基づいて書かれている。	38	0	3.42	4	1.177	1	5	0.876	<.001
		次の課題を明らかにすることが出来ている。	38	0	4.18	4	0.955	1	5	0.75	<.001
		AI作成	書き方	「常体（文末に「だ、である」を使う）」で文章が書かれている。	38	0	4.34	5	1.169	1	5
いつ、どこで、だれが、なにを、どうしたのか、具体的に客観的事実が書かれている。	38			0	4.34	5	0.938	2	5	0.712	<.001
要点が簡潔に書かれている。	38			0	4.42	5	0.826	2	5	0.702	<.001
分かりやすい表現で書かれている。	38			0	4.32	5	0.873	2	5	0.755	<.001
適切な専門用語や略語が使われている。	38			0	4.16	4	0.886	2	5	0.814	<.001
1文の長さを35字以内になっている。	38			0	3.95	4	1.207	1	5	0.803	<.001
必要時、利用者の言葉がそのまま書かれている。	38			0	2.76	3	1.515	1	5	0.859	<.001
他から得た情報は情報源も書かれている。	38			0	3.34	3	1.279	1	5	0.898	0.002
誤字・脱字がない。	38			0	4.32	5	1.068	1	5	0.669	<.001
不適切な表現がない。	38			0	4.21	4	0.991	1	5	0.757	<.001
文章の内容	プライバシーへ配慮して、名前のアルファベット表記など匿名化されて書かれている。		38	0	4	5	1.356	1	5	0.737	<.001
	その日の実習目標に照らして、その場面を詳細に記述し、考察がされている。		38	0	3.79	4	1.143	1	5	0.852	<.001
	体験した内容を踏まえて、その日の実習目標がどの程度達成されたのか、達成されなかったのかが書かれている。		38	0	3.89	4	1.203	1	5	0.802	<.001
	上記の達成された・されていないと判断した理由について書かれている。		38	0	4	4	1.09	1	5	0.813	<.001
	判断の根拠となる客観的な事実が書かれている。		38	0	3.95	4	1.184	1	5	0.803	<.001
	要因を分析して具体的に書かれている。		38	0	4.03	4	1.026	2	5	0.765	<.001
	知識や技術など、専門性に基づいて考察をすることが出来ている。		38	0	4.03	4	1.026	2	5	0.819	<.001
	自立支援、自己選択・自己決定等の視点に基づいて書かれている。		38	0	3.89	4	1.158	1	5	0.832	<.001
	次の課題を明らかにすることが出来ている。		38	0	4.16	4	1.027	1	5	0.758	<.001

表 6 書き方に関する評価項目における学生作成と AI 作成の比較

	N	平均値			ウィルコクソン順位	
		学生作成	AI 作成	差	P値	差
他から得た情報は情報源も書かれている。	38	4.16	3.34	0.82	< .001	1.5
必要時、利用者の言葉がそのまま書かれている。	38	3.00	2.76	0.24	0.323	0.5
「常体（文末に「だ、である」を使う）」で文章が書かれている。	38	4.32	4.34	-0.02	0.807	-7.28e-5
不適切な表現がない。	38	4.13	4.21	-0.08	0.662	-7.88e-5
誤字・脱字がない。	38	4.13	4.32	-0.19	0.264	-1
いつ、どこで、だれが、なにを、どうしたのか、具体的に客観的事実が書かれている。	38	4.11	4.34	-0.23	0.398	-0.5
分かりやすい表現で書かれている。	38	4.03	4.32	-0.29	0.198	-0.5
適切な専門用語や略語が使われている。	38	3.55	4.16	-0.61	0.003	-1
要点が簡潔に書かれている。	38	3.74	4.42	-0.68	0.005	-1.5
1文の長さを35字以内になっている。	38	3.16	3.95	-0.79	< .001	-1.5

※ P 値はウィルコクソン符号付順位和検定 (Wilcoxon signed-ranks test) による。

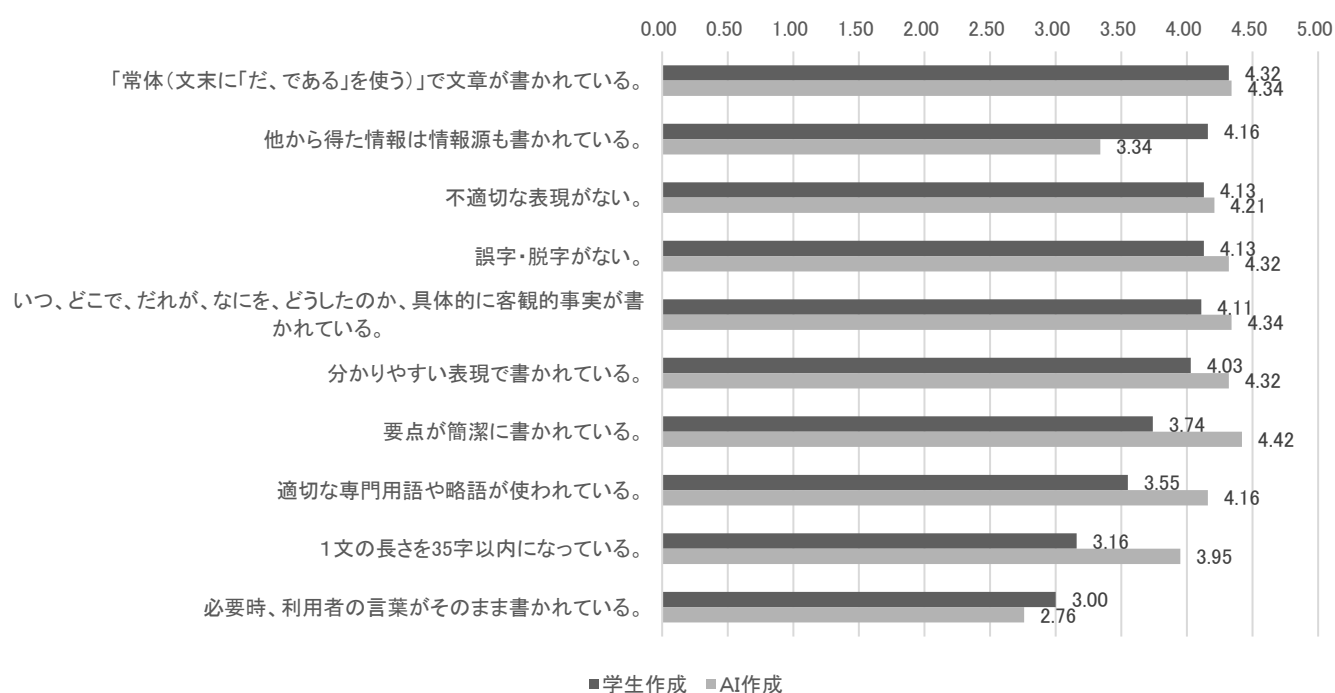


図 1 書き方に関する評価項目における学生作成と AI 作成の比較

学生作成と AI 作成で書き方に関する評価項目ごとの平均値を、ウィルコクソン符号付順位和検定 (Wilcoxon signed-ranks test) で学生作成と AI 作成の 2 群の各評価項目の中央値に差があるかどうかを検定して、帰無仮説である「学生作成と AI 作成では評価項目の点数に差は無い」を検証した。その結果 (表 6, 図 1), 学生作成の方が有意に最も高かった評価項目は「他から得た情報も書かれている」であった (差: 0.82, $P < 0.001$)。逆に AI 作成の方が有意に高かった (マイ

ナス差として表記) 評価項目は「1 文の長さが 35 字以内になっている」 (差 = -0.07, $P < 0.001$), 「要点が簡潔に書かれている」 (差: -0.68, $P = 0.005$), 「適切な専門用語や略語が使われている」 (差: -0.61, $P = 0.003$) の 3 項目であった。それ以外の項目についてはその差に有意性を認められなかった。

書き方に関する評価項目の比較から言えることは、①全体を通して AI 作成の方が評価が高い、②文字数や専門用語は AI 作成の方が評価が高い、③学生作成

の方が他から得られた情報まで書くことが出来ているの3点である。形式や専門知識といった点はAIの方が得意であるが、他から得た情報というその臨床場面を経験しているからこそ記述するが可能となる項目については、その臨床場面を経験した学生が自ら文章を作成するからこそ記述されるものであると言える。

次に考察内容に関する評価項目について、ウィルコ

クソン符号付順位和検定 (Wilcoxon signed-ranks test) で学生作成とAI作成の2群の各評価項目の中央値に差があるかどうかを検定して、帰無仮説である「学生作成とAI作成では評価項目の点数に差は無い」を検証した。その結果 (表7, 図2), 学生作成の方が評価が有意に高い項目は無かった。それに対してAI作成の方が有意に高かった (表ではマイナス表記) のは

表 7 考察内容に関する評価項目における学生作成とAI作成の比較

	N	平均値			ウィルコクソン順位	
		学生作成	AI作成	差	P値	差
プライバシーへ配慮して、名前のアルファベット表記など匿名化されて書かれている。	38	4.37	4.00	0.37	0.067	1
次の課題を明らかにすることが出来ている。	38	4.18	4.16	0.02	0.978	0.000053
その日の実習目標に照らして、その場면을詳細に記述し、考察がされている。	38	3.74	3.79	-0.05	0.59	-3.89e-5
判断の根拠となる客観的な事実が書かれている。	38	3.87	3.95	-0.08	0.609	-8.40e-5
要因を分析して具体的に書かれている。	38	3.95	4.03	-0.08	0.538	-1.37e-5
知識や技術など、専門性に基づいて考察をすることが出来ている。	38	3.71	4.03	-0.32	0.036	-1
体験した内容を踏まえて、その日の実習目標がどの程度達成されたのか、達成されなかったのかが書かれている。	38	3.53	3.89	-0.36	0.052	-1
自立支援、自己選択・自己決定等の視点に基づいて書かれている。	38	3.42	3.89	-0.47	0.028	-1
上記の達成された・されていないと判断した理由について書かれている。	38	3.45	4.00	-0.55	0.004	-1

※ P値はウィルコクソン符号付順位和検定 (Wilcoxon signed-ranks test) による。

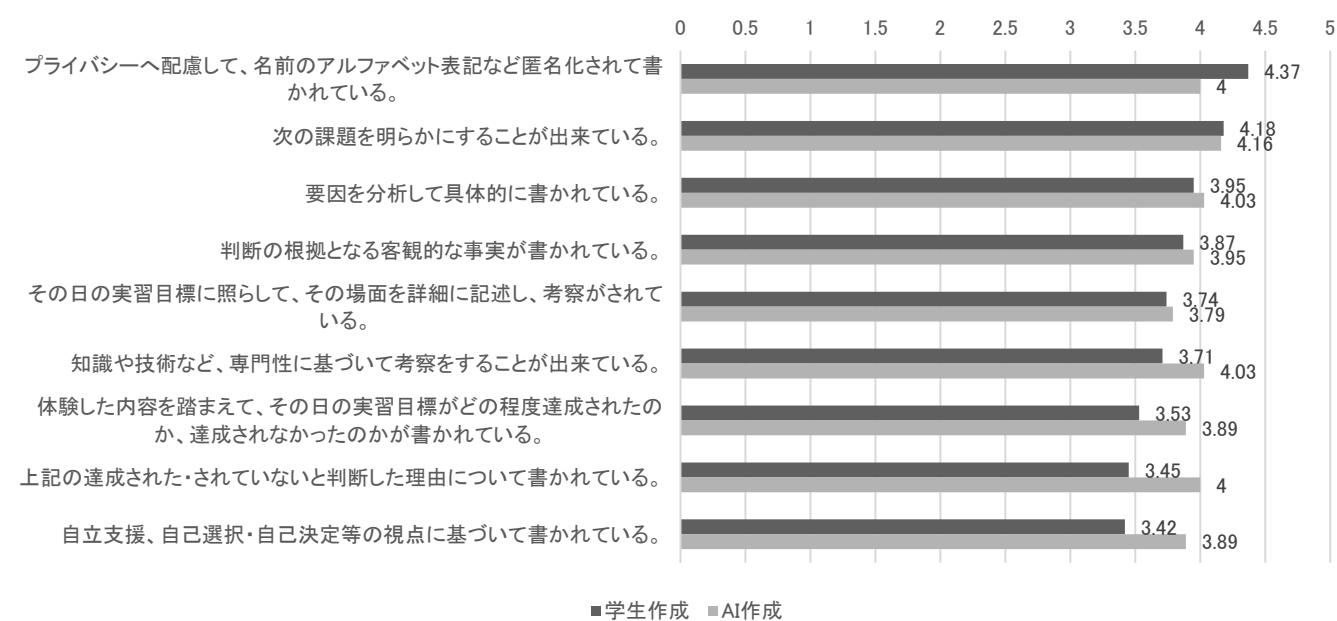


図 2 考察内容に関する評価項目における学生作成とAI作成の比較

「上記（実習目標）の達成された・されていないと判断した理由について書かれている」（差：-0.55, $P=0.004$ ）, 「自立支援・自己選択・自己決定等の視点に基づいて書かれている」（差：-0.47, $P=0.028$ ）, 「知識や技術など、専門性に基づいて考察をすることができている」（差：-0.32, $P=0.036$ ）の3項目であった。それ以外の項目は帰無仮説を棄却できなかった。これらから言えることは、考察内容においても AI 作成の方が全体的にも評価は高かったということである。

5. 結果の考察

以上の研究結果について、専門職として求められる判断の根拠を書くという点や自立支援・自己選択・自己決定等の福祉における基本的視点に基づいて書くことが出来ているかどうかでは、AIの方が評価が高かったという点は注目に値する。また実際に体験した学生と、実際には体験をしていない AI の間に具体的な記述に関する項目でも差が認められなかったという点は、AI が作成した文章でも実習という臨床体験をした者と同等レベルの文章を作成することが可能であるということを示していると言える。

この点について掘り下げると、実習記録を指導者がそこに書かれている考察文章を「誰が書いた文章か」見抜くことは難しいのではないかと考えられる。AIを活用することで、実習生としては短期的には実習日誌の考察文章を容易に作成することができる。自らの経験をどのように考察文章へと書き上げていけば良いかという「見本となる文章」から、どのように考察文章を書けば良いのかというコツを得るという点では大変有効であろう。しかしそれは、中長期的には専門性に基づいた考察や対象者の機微への気付きなど、実習経験の中で獲得する学びを得る機会を逸するとも考えられる。そうすると、実習記録とは何か、さらには実習で学生は何を学ぶべきかという根本的な教育目標の再考と、AI が活用されることをどのように織り込んでおくのが、次に検討すべき課題であると言える。

6. おわりに

今後の研究の展開としては、学生が作成した文章が AI を上回ることはあるのか、あるとした場合、何日目

あたりで上回るのかを調査することが考えられる。併せて、AI に実習日誌データとソーシャルワークの専門知識について学習させた場合、どのような考察文を生成するのかを実験するテーマも考えられる。ソーシャルワーク教育における、文章生成 AI の活用に関する研究の端緒につくことはできたと言える。

謝辞

本研究は JSPS 科研費 19K02977 の助成を受けたものである。考察文章の提供に協力してくれた学生に、この場を借りて御礼を申し上げる。

参考文献

- (1) 坂本毅啓, 佐藤貴之, 中原大介: “福祉職・保育者養成教育における ICT 活用への挑戦 - より深い学びと質の高い支援スキル獲得を目指して -”, 大学教育出版, 岡山 (2019)
- (2) ソーシャルワーク教育学校連盟編: “厚生労働省 令和 4 年度生活困窮者就労等準備支援事業費等補助金 (社会福祉推進事業分) 「平常時の社会福祉士養成課程における ICT 活用方法の検証に関する調査研究事業」実施報告書”, 日本ソーシャルワーク教育学校連盟, 東京 (2023)
- (3) 坂本毅啓, 佐藤貴之, 中原大介: “実習記録の ICT 活用に反対する意見の検討”, 地域ケアリング, Vol.23, No.9, pp.52-55 (2021)
- (4) 公益社団法人日本社会福祉士会編: “新版 社会福祉士実習指導者テキスト”, 中央法規, 東京 (2022)
- (5) 一番ヶ瀬康子監修, 木下安子: “介護福祉ハンドブック 介護実践記録の書き方”, 一橋出版, 東京 (2005)
- (6) 一般社団法人日本社会福祉士養成校協会編: “相談援助実習指導・現場実習 教員テキスト 第 2 版”, 中央法規, 東京 (2015)
- (7) 相澤譲治, 九十九綾子編相談援助実習 一養成校と実習先との連携のために一”, 電気書院, 東京 (2016)
- (8) 富川雅美: “よくわかる介護記録の書き方 第 5 版”, メヂカルフレンド社, 東京 (2017)
- (9) 杉本浩章・田中和彦: “実習生必携 ソーシャルワーク実習ノート [第 2 版]”, みらい, 岐阜 (2015)
- (10) 柳井久江: “4Steps エクセル統計【第 4 版】”, オーエムエス出版, 東京 (2015)

大学 1 年生の生成 AI 利活用に関するアンケート調査報告

森 祥寛

金沢大学学術メディア創成センター

Report on a Survey Regarding the Utilization of Generative AI among First-Year University Students

Mori Yoshihiro

Emerging Media Initiative, Kanazawa University

With the advent of 2023, the use of generative AI such as ChatGPT in education has gained attention. Therefore, we conducted a survey on the use of generative AI among the first-year students who entered Kanazawa University in April of the academic year 2023. This report presents the results of that survey.

キーワード: 生成 AI 利活用, 調査報告

1. はじめに

近年、情報技術はさまざまな分野で活用されている。2010 年代始めには ICT などが、2010 年代半ばからはデータサイエンスや DX, Society5.0 などの言葉で表わされる情報技術を元にした社会の変革が求められてきた。例えば、第 5 期科学技術基本計画（2016 年度から 2020 年度）⁽¹⁾には、データドリブンな社会を目指すことが明記されている。これらは、データがわれわれの日常生活のあらゆる面で増加し、それを理解し解釈する能力が必要とされていることから明らかである。このために、高等教育機関に対しては、いわゆる「データサイエンス教育の強化」が求められてきた。文部科学省では、高等教育における「数理・データサイエンス・AI 教育認定制度⁽²⁾」を実施するなど、さまざまな施策がとられている。

金沢大学では、2000 年代半ばより、大学全体で IT を教育に活用していくような取組が始まりだし、現代 GP や特色 GP のような文科省の補助事業によって後押しされてきた。併せて、最近では BYOD と呼ばれるような、入学時にパソコンやタブレット端末などを新入学生に準備させ、授業などで使用する取組も進められてきた⁽³⁾。同時に ICT を活用していくために必要な知識やリテラシーを学ぶ教育も進められた。このよう

な流れは、2010 年代後半になると、初等中等教育でも進められるようになり、「教育の情報化の推進⁽⁴⁾」として、2019 年 6 月 28 日に教育の情報化の推進に関する法律が公布、施行されることで、GIGA スクール構想が加速し、2021 年 4 月からは、新しい学習指導要領に則った高等学校情報科「情報 1」が実施された。特に情報 1 は、大学入試共通テストでの情報の実施と併せて、初等中等教育における情報教育に大きな影響を与えている。

そのような中で、2022 年 11 月に登場したのが、OpenAI の GTP（大規模言語モデル）^(5, 6)を使った ChatGPT が一般リリースされた。これは大量のテキストデータを使ってトレーニングされた自然言語処理のモデルであり、チャット形式で質問を書くと AI が回答をしてくれるサービスである。このサービスによって、2023 年始めに「生成 AI 利活用」という新しい変革が訪れた。この変革は、教育の現場にも影響をあたえた。生成 AI によって、学習者の学習およびその活動の在り方が大きく変わる可能性がでてきたのである。

メリットとして捉えるなら、ChatGPT のような大規模言語モデルは、学生が自己主導的に学習を進めるための強力なツールとなることがあげられる。学生は、自分のペースで学習を進め、深く理解するために必要

な質問を AI に投げかけることが可能となった。また、教員は、AI が学生の質問に対応することで、より高度な教育活動に集中することが可能となった。さらに、ChatGPT のような生成 AI は、上手に使える、学習者が新たな知識を探索し、理解を深めるための新しい方法を提供できる。例えば、学生は、その時点における自分の理解度をテストするために AI に質問を投げかけることができ、それに対して AI は学生が理解を深めるための新たな視点や情報を提供することが可能である。デメリットとして捉えるなら、教員が出した課題を AI に質問として投げかけることで、学生が自らの学習を経ることなく回答を得ることが可能となることがあげられる。他にも AI を使用することで、データ流出やプライバシーの侵害、AI がハルシネーション（事実に反することやデタラメを出力する現象）を引き起こす可能性がある。

これらを踏まえて、本研究では、2023 年 4 月に金沢大学に入学した 1 年生に対して生成 AI 利活用に関するアンケート調査を行い、その利用状況、認識などについて評価した。

2. 調査概要

2.1 調査方法と対象

本研究は、金沢大学の LMS (WebClass⁽⁷⁾) を使用して、2023 年度に入学した全 1 年生を対象にしたオンライン上のアンケート調査として実施した。実施期間は、生成 AI について解説をしたデータサイエンス基礎第 3 回目の講義（4 月 20 日から 26 日に実施）終了後、ゴールデンウィークを経て、5 月 12 日から 5 月 19 日までの 1 週間である。この調査期間設定は、講義を受けた上で、実際に学生が生成 AI を使用するに足る十分な時間をとるためのものである。

2.2 調査項目

調査項目として、全 14 問用意した。そのうち 12 問は多肢選択式の設問で、記述式の設問は最後の 2 問のみである。設問では、生成 AI について解説したデータサイエンス基礎第 3 回目の講義を起点として、その前後で生成 AI について使用したか否かを聞き、その上で生成 AI に対する疑問や懸念、生成 AI 利用における注目点、学校教育でどのように利用されることになる

と予測するか、利用目的、活用場面などについて聞いている。

なお、各調査項目については、金沢大学固有の設問となるデータサイエンス基礎に関連する部分を削除し、回答者属性（所属や学年、研究内容）を加えたものを Google Form (<https://forms.gle/ZjPrtpVcUeih5Smr5>) に公開している。

2.3 回答数など

本調査を実施した結果、表 1 のような回答数を得た。

表 1 回答者属性と回答者数

回答者所属	学生数	回答者数	回答率
文系	692	168	24.3%
理系	609	152	25.0%
医系	390	121	31.0%
文理融合	89	34	38.2%
合計	1,780	475	26.7%

あくまでも調査協力の形を取っているため、1 年生全体の 26.7%からの回答であったが、本件調査においては十分と考えている。一方で金沢大学では、20 の学類（他大学における学部相当）があり、学類毎の分析では、回答数が小さくなる学類もあるため、これをおこなわなかった。代わりに、学類をいわゆる文系、理系、医系、その他という形で区分し、これを本調査においては「回答者属性」とした。金沢大学では学類の上部区分として学域が設定されており、人文社会学域を文系、理工学域を理系、医薬保健学域を医系、融合学域を文理融合に対応させた。

3. 調査結果と考察

3.1 生成 AI の使用状況

本調査の結果、金沢大学 2023 年度 1 年生が、データサイエンス基礎第 3 回講義前までの段階で、生成 AI を使っていたか否かについてまとめたものが図 1 である。2023 年 4 月半ばの段階では、属性に関わりなく 8 割方の 1 年生が生成 AI について知っていた。一方で、使ったことがある学生は、3 割程度に留まっていた。

調査では、講義後に生成 AI を使用してみたかも聞いており、その結果、生成 AI の使用が、文系 50%、理系 42.8%、医系 52.9%、その他 70.6%となった。こ

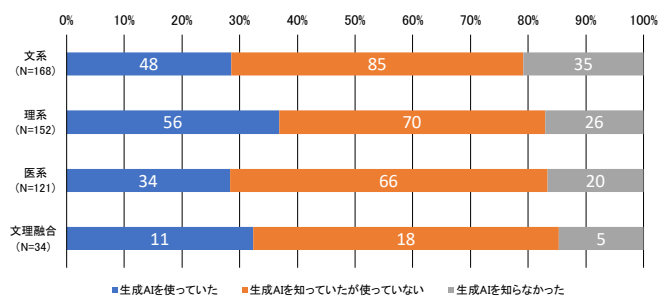


図 1 講義前に生成 AI を使っていたか

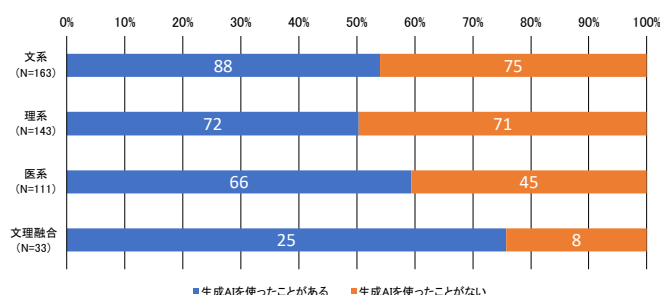


図 2 講義後に生成 AI を使ってみたか

これらの結果を踏まえて、分析において、学生を「生成 AI を使ったことがある」グループと「使ったことがない」グループに分類して、分析を行うこととした。なおこれらの設問への未回答者については、分析から除いたため、回答者数は文系 (N=163)、理系 (N=143)、医系 (N=111)、その他 (N=33) となっている (N は

回答者数を示す)。

3.2 生成 AI への疑問や懸念、利用目的

図 3 は、生成 AI に対する疑問や懸念に対する回答である。回答者属性と生成 AI の使用の可否を併せて掲載している。学生がもつ生成 AI への疑問点として、回答の正確性や質保証、生成 AI の利用実績が高く上げられており、懸念として、生成 AI を利用することで不正を疑われるのではないかとというのが高くなっていた。多くの回答において、属性による違いを判断するのは困難であった。今後、より深い分析をすることで何らかの違いがでてくることを期待したい。生成 AI の使用の有無では、使ったことがない学生の方が、使い方が難しそう、活用場所が分からないなどが多くなっており、生成 AI を使ってみることで、その利活用に対するハードルが下がることが見て取れる。コストや学習効果に対する懸念についても、使ってみることで下がる傾向が見られ、逆に不正を疑われるのではないかとという懸念に対して、生成 AI を使ったことがある学生の方が高くなる点は、学生の心理的傾向を示す一端であろう。

図 4 は、生成 AI の利用目的をまとめたものである。

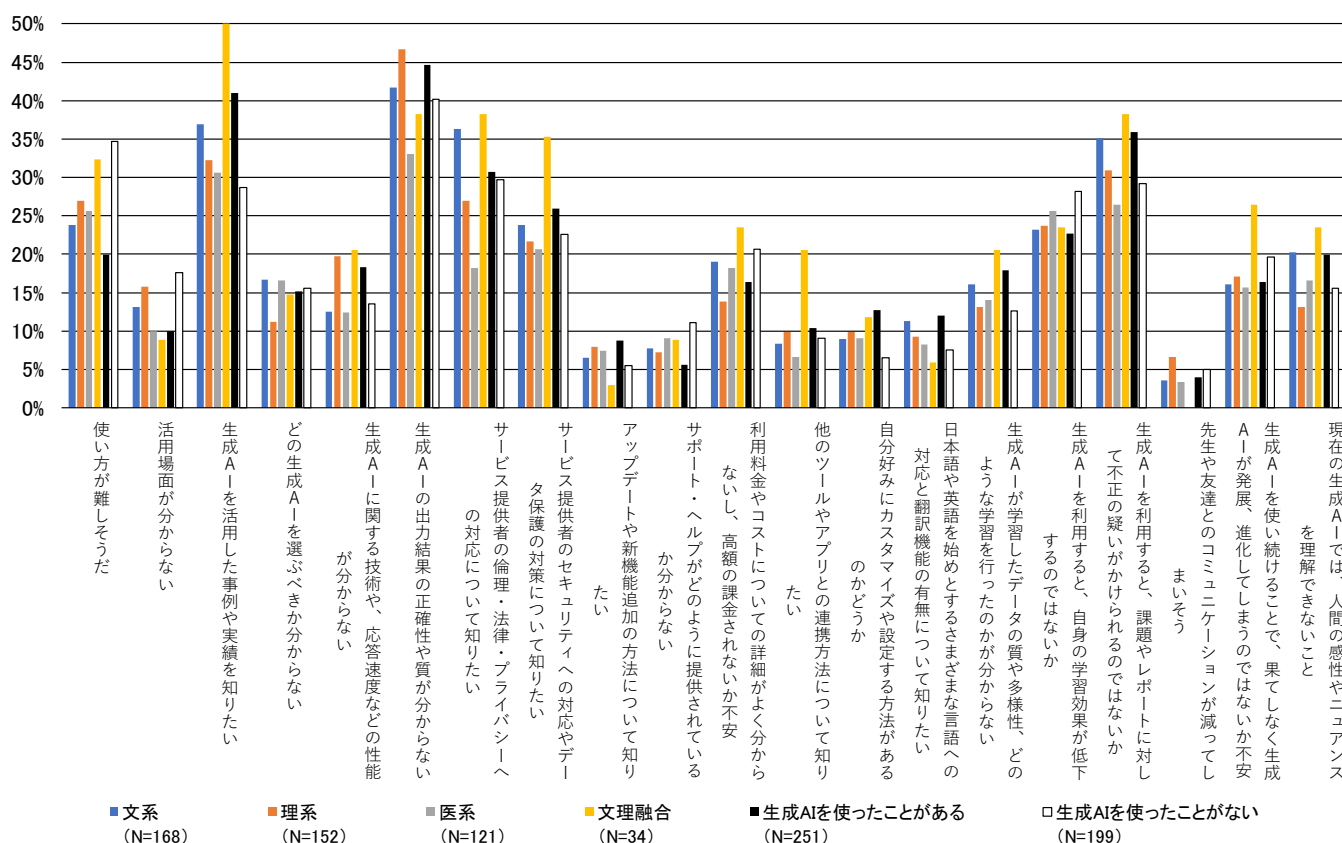


図 3 生成 AI に対する疑問や懸念

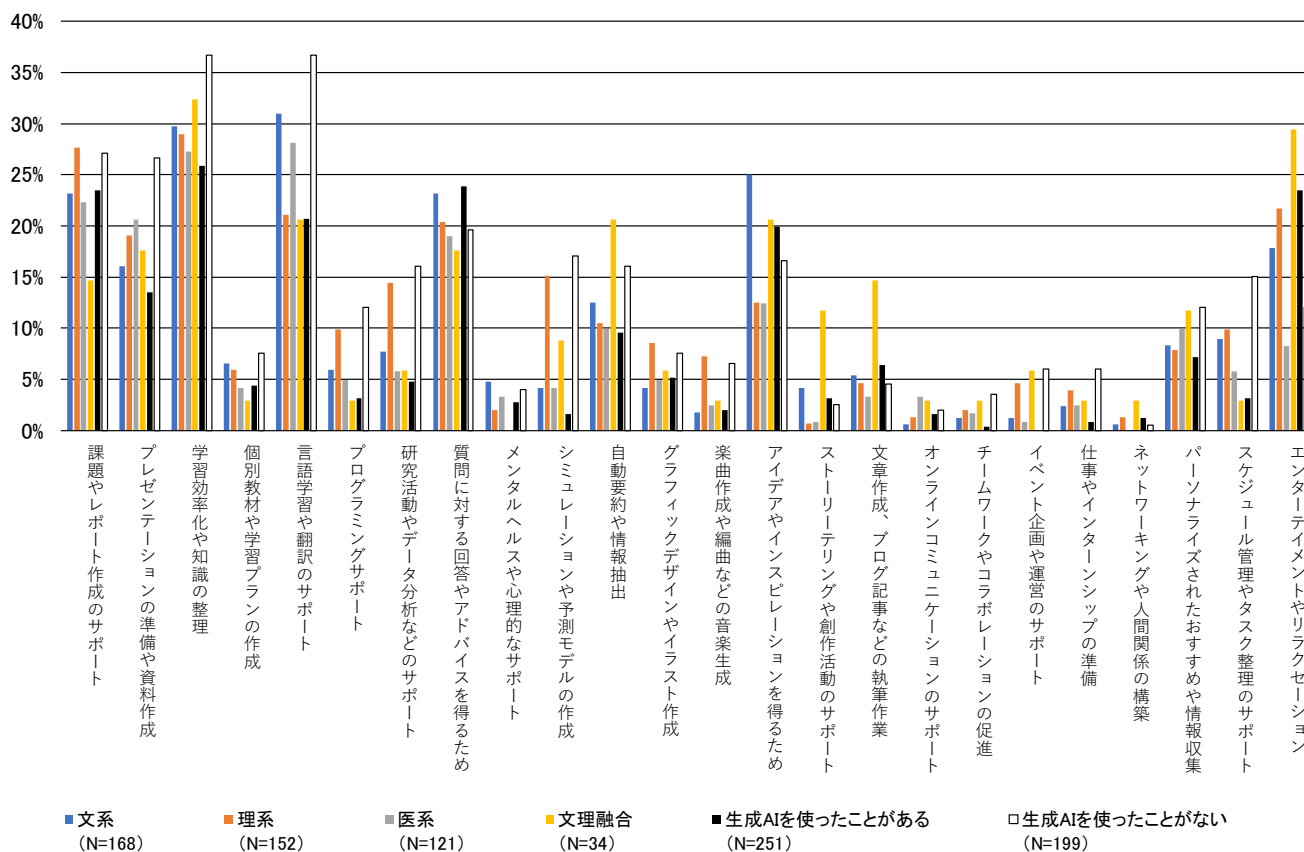


図 4 生成 AI の利用目的

生成 AI を使用していない場合は、利用したい目的を回答させた。生成 AI の利用目的は、回答者属性によってある程度の変化が存在することが見て取れる。それが良く見て取れるのが、プログラミングサポート、データ分析、シミュレーションや予測モデルの作成である。これについては、理系属性の学生が目的として選ぶことが突出しており、学生が行う学習内容への生成 AI への期待が見て取れる。同時に、これらの項目を選択しているのが、生成 AI を使ったことがない学生が選択していることも注目すべきであろう。学生は、実際に生成 AI を使用することで、2023 年 5 月現在の生成 AI が、これらの項目に対して期待しづらいことを

学生自身が理解すると考えられる。

3.3 生成 AI への期待

本調査では、「生成 AI を利用することで、文章作成が効率化されるという意見に、あなたはどの程度同意しますか。」についても尋ねた。その結果が、図 5 である。調査で尋ねているのが、文章作成の効率化であり、その是非ではないため、学生の意見としては、不同意より同意に傾いていることが見て取れる。それらは、回答者属性による違いは見られなかった。一方で、生成 AI を使ったことがあるか否かで見てみると、生成 AI を使ったことがある学生の方がより同意をする傾

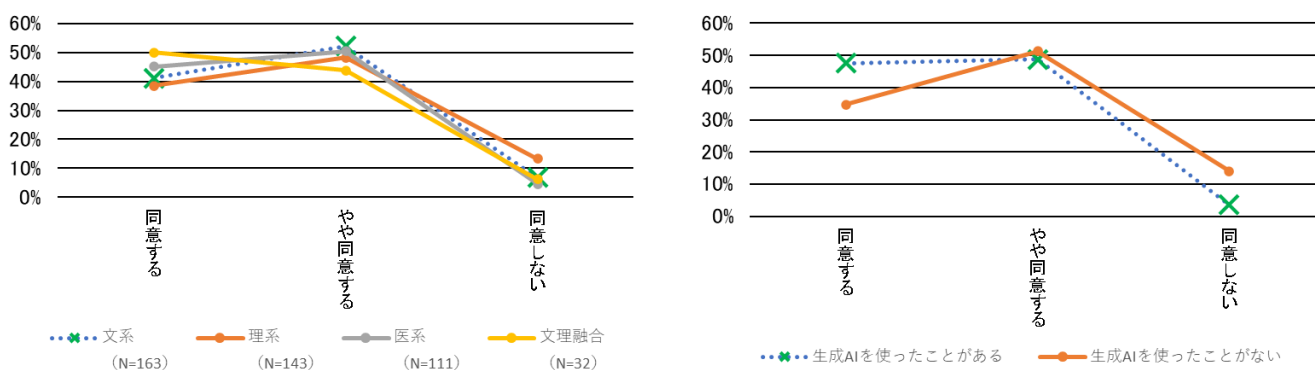


図 1 生成 AI 利用による文章作成効率化への同意・不同意

ここで尋ねている生成 AI は、GTP シリーズに代表される自然言語処理のためのディープラーニングモデルの 1 つを前提としている。実際に ChatGPT や BingAI などを使用した学生にとっては、文章作成が効率化されるということ、実感を持った回答と捉えられる。一方で、生成 AI を使用していない学生は、メディアなどから得た生成 AI に関する情報で AI による文章作成について判断していると考えられ、その結果が多少同意しないが多くなったと捉えられる。この辺りについては、より詳細な調査が必要であろう。

これは、生成 AI を使用した学生 (N=251) への質問となるが、生成 AI を使用した上での生成 AI への不満点を尋ねた。その結果が図 6 である。

3.5 使用した／使用したい生成 AI

表 2 使用した／使用したい生成 AI

その多くが、ChatGPT を選択している。これは講義における説明の中心が ChatGPT であったこと、メディアなどで発信される情報が ChatGPT を中心としたものであったことが原因といえる。しかしながら、現状では、「生成 AI=ChatGP」という形が揺るがしがたいということを示した、一事例といえる。

A bar chart titled "不満の理由" (Reasons for Dissatisfaction) showing the percentage of respondents who cited specific issues as reasons for being dissatisfied with video quality. The Y-axis represents the percentage from 0% to 70%. The X-axis lists 18 different reasons. The bars are blue.

不満の理由	割合 (%)
テンプレートやサンプルの種類・質に不満	4%
フィードバック・改善提案の反映が遅生感あり、またはされない	65%
生成AIによるような不自然な発音・明瞭性の不足（どのようなテラフォーマットで学習しているかわからない）	9%
音声合成・音声入力機能の品質や対応言語に不満	4%
画像や動画との連携機能の不足（フレイクアップルの取り込みや編集がオンラインでの利用が制限される）	14%
更新・バージョンアップの頻度や内容の不透明	3%
デバイス対応の不足（スマートフォンやタブレットでの利用が難しい）	8%
コラレーション機能の不足（複数人での共同作業が難しい）	3%
生成結果の整理・導き出しに関する懸念（不適切な内容や偏見が含まれている）	4%
言語対応の不足（利用したい言語が対応していない）	8%
ユーザーインターフェイス（UI）や操作性の問題（画面が見づらい、操作が難しすぎる）	11%
プライバシー・セキュリティの懸念（個人情報保護と安全対策）	8%
サポート・教育資料の不足（使い方やヘルプが充実してない）	4%
カスタマイズ性（個別のニーズに合わせて調整できるかどうか）	13%
利便性・使いやすさ（操作性やアクセスコスト（利用料金やコストパフォーマンス））	9%
スピード（生成にかかる時間が早いかどうか）	65%
精度（生成される内容の正確さと適切さ）	4%

合、ChatGPT の表記ミスなども多々存在した。表 2 中の括弧書きの数字が、間違った表記数である。大文字小文字の混在の他、Chat と GPT の間にスペースが入るか否か、Chat をチャットと表記するなど、間違いではない記述の他、GPT が GTP、GDP、GBT など新しいワード故の間違いも存在していた。

今回、生成 AI の活用について、2023 年度に金沢大学に入学した 1 年生に対して、アンケート調査を行った結果を報告した。

—55—

謝辞

本研究は、金沢電子出版株式会社との共同研究により推進された。関係各位のご厚意ご高配に、深く感謝する。

参 考 文 献

- (1) 内閣府・科学技術政策：科学技術基本計画，
<https://www8.cao.go.jp/cstp¥¥/kihonkeikaku/index5.html>(2023年6月14日確認)
- (2) 文部科学省：数理・データサイエンス・AI教育認定制度，
https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00001.htm(2023年6月14日確認)
- (3) 森祥寛，佐藤正英，大野浩之，笠原禎也，井町智彦，高田良宏，東昭孝，二木恵，NAKASANCHAWANAT, "金沢大学における携帯型パソコン必携化に関する12年間の取組", 学術情報処理研究 23(1), 29-42, 2019
- (4) “教育の情報化の推進”，
http://www.mext.go.jp/a_menu/shotou/zyouhou/index.htm(2023年6月14日確認)
- (5) A. Radford and J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever: "Language Models are Unsupervised Multitask Learners", OpenAI blog, 1(8):9.
- (6) T. B. Brown and B. Mann, et al.: "Language models are few-shot learners", Proceedings of the 34th International Conference on Neural Information Processing Systems, 2020
- (7) 日本データパシフィック株式会社：WebClass, <https://webclass.jp/>(2023年6月14日確認)

学習支援ロボットのロールによるセルフレビューへの影響

佐田 竣祐^{*1}, 柏原 昭博^{*1}

^{*1} 電気通信大学

Effects of Robot Roles on Presentation Self-Review

Shunsuke SADA^{*1}, Akihiro KASHIHARA^{*1}

^{*1} University of Electro Communication

人間同士のコミュニケーションでは、相手の役割や立場によって、コミュニケーションに認知的なバイアス加わり、思考や情動、接し方が変化する現象がある。ソーシャルロボットが他のメディアと比較して擬人化傾向を持つことに着目すると、学習支援ロボット相手でも、そのロールによって人間のコミュニケーションにバイアス加わり、思考や情動・接し方が変化すると考えられる。この現象は学習者の学びに影響を与える可能性があるが、学習支援ロボットのロールに着目し、その影響を検証する研究は十分に行われていない。本研究では、「学習者のプレゼンテーションをロボットが再現し、それを学習者自身がレビューする」文脈で、学習支援ロボットのロールによる学習者への影響を検証した。その結果、「学習者のピア」のロールを担うロボットでのセルフレビューが、学習者の心理的抵抗感を軽減し、自分自身のプレゼンテーションの改善すべき点への気づきを促す可能性が示唆された。

キーワード: 学習支援ロボット, バイアス, プレゼンテーション, セルフレビュー支援

1. はじめに

近年、学習者とのコミュニケーションを伴う学習パートナーとして、ロボットを活用する学習支援の研究が進められている。筆者らも、人間の教師の代わりに講義を行うロボット[1], 学習者と英文読み合いを行うロボット[2], 学習者のプレゼンテーションを再現することでセルフレビューを支援するロボットシステム[3]を開発してきた。これらの研究からは、ロボットを学習パートナーとして用いることで、学習者のエンゲージメント・客観視の促進、心理的抵抗感の軽減につながる事が示唆されている。

人間同士のコミュニケーションでは、コミュニケーション相手の役割・立場によって認知的バイアス加わり、思考や情動・接し方が変化する現象がある[4][5]。本研究では、このコミュニケーション相手の「役割・立場」をロールと呼ぶ。ソーシャルロボットの関連研究では、ロボットは他のメディアと比較して身体性を有し、擬人化傾向を持つという特性があり、あたかも人間のように扱われることが示唆されている[6]。この

特性を考慮すると、学習支援ロボットのロールが変われば、学習者のコミュニケーションにかかるバイアスも変化し、学習者の学びに影響を与えると考えられる。しかし、学習支援ロボットのロールによる影響については、これまで十分に研究されていない。そこで、本研究では学習者の学びにおけるエンゲージメント、心理的抵抗感、客観視に着目し、これらに対して学習支援ロボットのロールが与える影響を検証することを目的としている。

筆者らが開発している講義ロボット[1]では「教師」のロール、英文読み合いロボット[2]では「学習者のピア」のロール、プレゼンテーションセルフレビュー支援ロボット[3]には「学習者自身」のロールを与え、それぞれ学びの支援を実施している。本稿では、このうち、プレゼンテーションセルフレビュー支援について、ロボットのロールを変化させた場合の学習者のセルフレビューへの影響を検証した実験について論じる。ここでは、「ロボットが学習者のプレゼンテーションを再現し、学習者自身がそれをレビューする」という文脈で実験を行った。2 台のプレゼンテーションセルフレ

ビュー支援ロボットを準備し、それぞれ「学習者自身」と「ピア」という異なるロールを設定し、各ロールを担うロボットのプレゼンテーション再現に対する学習者のレビューを比較することで、ロボットロールの影響を検証した。

その結果、ピアのロールでのセルフレビューが、学習者の心理的抵抗感を軽減し、自分自身のプレゼンテーションの改善すべき点への気づきを促す可能性が示唆された。

2. コミュニケーションにおける認知的バイアス

2.1 コミュニケーション相手のロールが与える影響

人間同士のコミュニケーションでは、相手の役割や立場によって、コミュニケーションに認知的なバイアス加わり、思考や情動、接し方が変化する現象がある。関連研究[4]では、電子メールによるコミュニケーションにおいて、相手の立場が及ぼす影響を調査している。この研究では、相手の立場を「目上」、「匿名」、「同僚」、「友人」の4つとし、どの相手に返信メールを送るかによってメール内容や書き方、書くときの感情に影響を与えることが示唆された。また、関連研究[5]では、チューターがある題材について「学生」に教える場合と、「自分自身」に教える場合に生じる違いを検証している。その結果、相手が「学生」の場合と「自分自身」の場合で説明の仕方が異なり、「自分自身」に説明する方が知識構築につながる発言が多くなることが示唆された。本研究では、こうした「同僚」や「学生」のような相手の立場・役割を「ロール」と呼んでいる。

以上のようなコミュニケーション相手のロールが認知に与える影響は、人間同士だけではなく、ロボット相手でも生じるだろうか？筆者らは、ソーシャルロボットが擬人化傾向を持つこと[6]に着目し、ロボット相

手でも、そのロールによって人間のコミュニケーションにバイアス加わり、思考や情動・接し方が変化するであろうと考えている。文献[7]でも、ロボットに「男性」「女性」という性別を与え、それぞれのロボットに被験者から説明させ、その内容を比較する実験を行っている。ロボットの性別は、声や見た目を変えて表現している。実験の結果、ロボットに与える性別によって説明の詳しさが変化的ことが示唆された。この結果から、目的に応じた性別を与えてロボットを設計する重要性が示唆されている[7]。また、コミュニケーション相手であるロボットに「ニューヨーク」、「香港」という出身地の設定を与え、それぞれのロボットが有する知識に対して、被験者がどのような認識を持つかを比較する実験が行われている[8]。ロボットの「出身地」は、ロボットが作られた場所と話す言語で表現している。実験の結果、ロボットの「出身地」によって、ロボットが持つ知識に対する認識が変化的ことが示唆された。これは、ロボットの出身地が、ロボットの知識を推定する際に認知バイアス加えることを示唆している。

2.2 学習支援への応用

人間同士のコミュニケーションにて、相手のロールによって生じる認知的バイアスは、コミュニケーションを伴った学習においても生じると考えられる。文献[9]では、教員の性別が女子学生の履修行動や専攻選択に与える影響を検証した。初めて履修する科目に注目し、データセットを用いた分析を行ったところ、同性（女性）の教員による科目が、学生の学問への興味にプラスの影響を与えることが示唆された。

以上の研究から見られるのは、教員の性別によるバイアスの違いであるが、「教師」と「学生」のようなロールでも、バイアスに違いが生じると考えられる。図1に示すように、同じ内容（「AはBである」）を学習



図1 「教師」と「学生」のロールによる影響

者に教授していても、学習者にかかるバイアスは、相手が教師か学生かによって変化すると考えられる。この際、支援の文脈に応じて、適切な学習相手のロールは変わる。例えば、図 1 左側のように「情報の受容」を促す文脈では、学生が教えると学習者は教授内容の適切さを疑いながら聞いてしまう。一方、教師が教えると、学習者は教授内容を適切なものとして受容しようとすると考えられる。つまり、数学の公式や定義のように、受け入れて欲しい事柄について教授する場合は、教師のロールのほうが望ましい。一方、図 1 右側のように「情報の批評」を促す文脈では、教師が教えると学習者は教授内容に疑問も持たず受け入れてしまいがちになるが、学生が教えると教授内容が適切かどうか批評的な態度で聞こうとする傾向になると考えられる。つまり、学習者に教授内容について疑問を抱き質問してほしい場合は、学生ロールでの教授が望ましいといえる。

以上のような学習にかかる認知バイアスは、コミュニケーション相手が学習支援ロボットの場合でも生じると考えられる。例えば、関連研究[10]では指導ロボットの性別が課題を解くプロセスに及ぼす影響について検証を行った。実験の結果、指導ロボットに与える性別を変化させた場合、学習プロセスに対する被験者のエンゲージメントも変化することが示唆された。このように、ロボットのロールによって学習者にかかるバイアスが変わるのであれば、ロボットのロールを制御することで、より望ましい「思考や情動・接し方」を学習者から引き出すことができるようになる。しか

し、「学習支援ロボットのロールが学習者に与える影響」については、これまで十分に研究されておらず、ロールによるバイアスの変化が学習支援ロボット相手でも起こるのかは、十分に検証されていないのが現状である。

そこで、本研究では、「学習支援ロボットに与えられたロールが、学習者の学びに与える影響」について検証することを目的とする。本実験では、「ロボットが学習者のプレゼンテーションを再現し、学習者自身がそれをレビューする」という文脈にて、検証を行った。この学習文脈については、次章にて説明を行う。

3. プレゼンテーションセルフレビュー支援

3.1 プレゼンテーション

研究者にとってプレゼンテーションは、自分の研究成果を発表するための重要な研究活動である。プレゼンテーションは一般的に、発表資料、オーラル、非言語動作（ジェスチャー・パラ言語）から構成されている。上記 3 つの構成要素のうち、非言語動作はプレゼンテーションの内容を効果的に聴衆へ伝えるための重要な要素であり、動作の意図を持って適切に行うことが求められる[11]。例えば、聴衆の注意を喚起する「パラ言語」や「強調するジェスチャー」などの非言語動作や、聴衆の注意を誘導する「スライドへの顔向け」や「スライドへの指差し」などの非言語動作を適切に実施することが必要である。関連研究[3]では、プレゼンテーション動作の意図と非言語動作の関係をまとめたプレゼンテーション動作モデル（図 2）をデザイン

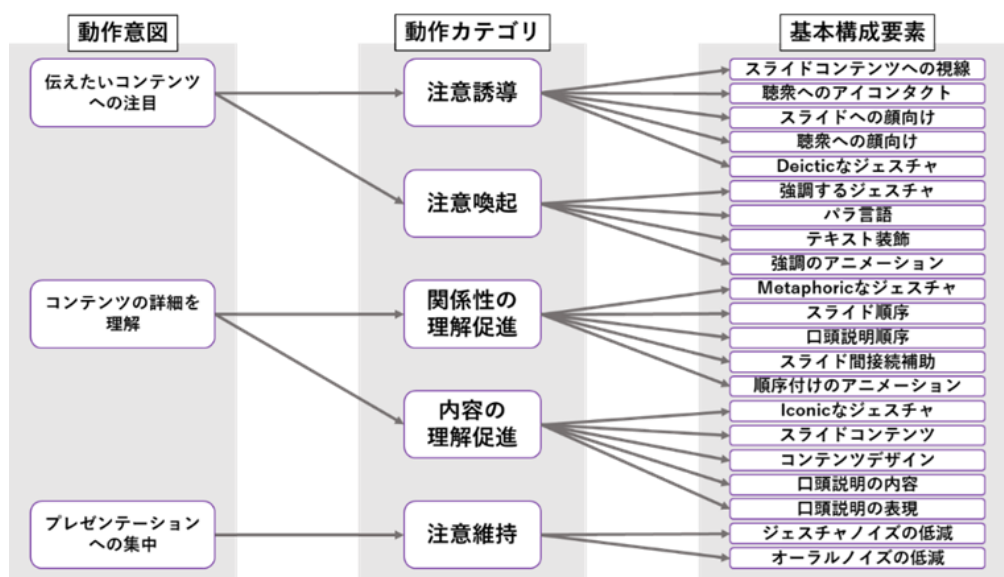


図 2 プレゼンテーション動作モデル

した．このモデルは3層から構成されており，動作意図を満たすためにどのような非言語動作が必要であることを示している．本研究では，被験者に動作意図を持って非言語動作を実施してもらうため，このモデルをもとに検証実験を行った．

3.2 プレゼンテーションセルフレビュー

プレゼンテーションを改善するためには，発表資料，オーラル，非言語動作の振り返り・見直しが必要である．その方法として，自分のプレゼンテーションを研究熟練者である他者に評価してもらうピアレビューと，自分のプレゼンテーションを動画撮影し，そのレビューを行うセルフレビュー[12]の二つが挙げられる．特に，セルフレビューは，他者の都合による制約を受けずに実施することができるが，動画内で自分の姿を見ることが自分の声を聞くことによる心理的抵抗感があり，また客観的にレビューを行うことは困難であるという問題点がある．また，何をレビューすべきかに関する知識が必要となるため，研究初学者にとって，適切にレビューを行うことは難しいという問題点もある．

3.3 モデルベースのセルフレビュー支援

以上のような問題点に対して，筆者らはプレゼンテーションのセルフレビューを，ロボットを用いて支援することを試みてきた．動画視聴にてセルフレビューを行う際に，自分の姿や音声に対して心理的抵抗感を感じ，適切なレビューを行うことができないという問題点に対し，学習者のプレゼンテーションを再現するロボットシステムを用いることを提案した[3]．その結果，学習者のセルフレビューに対する心理的抵抗感を軽減し，エンゲージメントを高め，改善点への気づきが促されることを確かめた．また，研究初学者が適切かつ十分にセルフレビューを行えないという問題点に対し，学習者のプレゼンテーションを診断し，2体のロボットによるフィードバックを行う手法を提案した[13]．その結果，改善点への気づきが促され，セルフレビューの改善につながる可能性を検証した．

3.4 ロボットロールの影響

筆者らは，これまでセルフレビュー支援において，学習者が「ロボットが自分のプレゼンテーションを再

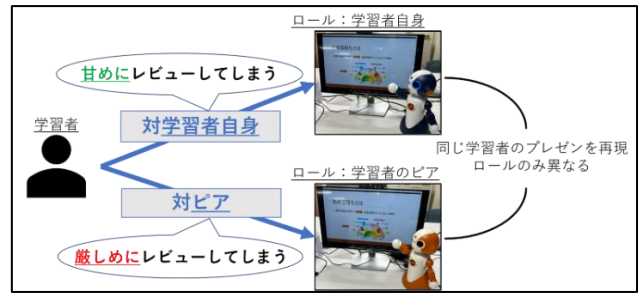


図3 セルフレビューでの検証実験

現している」と認識していることを前提としている．つまり，ロボットに「学習者自身」というロールを割り当てていた．本研究では，この「学習者自身」というロールを変化させた場合，学習者のセルフレビューにどのような影響が与えられるかを検証する．具体的には，図3のように，ロボットが「学習者自身」および「ピア」のロールで，学習者のプレゼンテーションを再現し，学習者にレビューを行わせることで，プレゼンテーションの客観視が促進されるかどうか，セルフレビューへの心理的抵抗感が減少するかどうか，エンゲージメントが高まるかどうかを確かめる．

4. ケーススタディ

4.1 実験計画

本実験では，2体の「プレゼンテーションを再現するロボット」に2種類のロール（「学習者自身」と「学習者のピア」）をそれぞれ設定し，各ロールにおける学習者のレビューを比較した．なお本実験は，関連研究[3]にて開発されたプレゼンテーションを再現するロボットシステムを用いて実施した．

本実験では，統制条件と実験条件を設定し，被験者内計画として実施した．統制条件は「学習者自身」のロールでのセルフレビュー，実験条件は「学習者のピア」のロールでのセルフレビューである．被験者は，各条件においてロボットが再現した（同じ）プレゼンテーションに対してセルフレビューを行った．ただし，統制条件では，ロボットのプレゼンが被験者自身のものであることを伝え，実験条件ではピアのものであることを伝えた．

被験者は，理工系の大学生・大学院生14名であった．被験者内実験では，セルフレビューの順番が結果に影響を与える可能性があるため，以下に示すとおり2群に分けた．

- A 群…1 回目のレビューは統制条件、2 回目のレビューは実験条件で実施する。
- B 群…1 回目のレビューは実験条件、2 回目のレビューは統制条件で実施する。

この実験では、恋声[14]にて同じ被験者の声からピッチの異なる 2 種類の声を作成した。群に関係なく 1 回目のセルフレビューでは高めの声を、2 回目のセルフレビューでは低めの声を使用した。さらに、この実験では 2 色の Sota[15]を使用した。こちらも群に関係なく 1 回目のセルフレビューでは青色の Sota を、2 回目のセルフレビューでは橙色の Sota を使用した。

また、被験者がプレゼンテーションで用いるスライドとオーラルの原稿は、予め実験者が用意した。

4.2 実験手順

まず被験者は、図 2 に示したプレゼンテーション動作モデルについて説明を受けた。次に被験者は、オーラルの原稿のみを用いて、10 分間プレゼンテーションの練習を行った。その後、事前に用意されたプレゼンテーション用スライドに対して、図 4 のように動作意図を付与した。次に、意図を満たす非言語動作を行えるように、動作意図を付与したスライドと原稿を用いて 10 分間プレゼンテーションの練習を行った。その後、非言語動作を意識しながらプレゼンテーションを行い、Kinect[16]を用いてモーションデータと音声を記録した。実験条件にて、被験者が「これは自分のプレゼンテーションを再現したものだ」と気づく可能性を考慮し、記録後に被験者は筆者らが準備した別の 10 分間のプレゼンテーション動画を視聴し、10 分間の休憩をとった。休憩後、群ごとに統制条件と実験条件でのプレゼンテーションのレビューを行った(図 5)。この際、被験者は Sota によるプレゼンテーションを見ながら、自分の非言語動作への改善点を書き出した。各レビューの制限時間は 15 分とした。各レビュー終了後、被験者は 5 段階のリッカート尺度を採用した事後アンケートに回答した。統制条件と実験条件で計 2 回のレビューを行った後、被験者は全体アンケートに回答した。これは、どちらのロールがより質問に当てはまるかを問うアンケートであった。

以上の実験計画に基づき、本実験では次の 3 つの仮説を設定した。

- 仮説 1…プレゼンテーションセルフレビュー支援ロボットのロールをピアにすることで、セルフレビューへのエンゲージメントが高まる。
- 仮説 2…プレゼンテーションセルフレビュー支援ロボットのロールをピアにすることで、セルフレビューに対する心理的抵抗感が軽減される。
- 仮説 3…プレゼンテーションセルフレビュー支援ロボットのロールをピアにすることで、客観視が促進される。

4.3 評価方法

本実験は 1 要因 2 水準の被験者内計画であるため、順序効果を考慮し 2 群に分けている。評価では、2 群での結果を条件ごとにまとめることで、両群間の差ではなく、両条件間の差について分析を行った。ピアのロールでのセルフレビューが、学習者のエンゲージメント、心理的抵抗感、客観視に変化をもたらすかどうかを、両条件における 2 つのアンケート結果を比較することで検証した。事後アンケートに対しては、各質問の結果について対応のある t 検定を行い、各質問の平均値に有意差があるかを検討した。さらに、Cohens'd を尺度として、平均値の差に対する効果量を検討した。また、被験者が書き出した非言語動作への



図 4 スライドへの動作意図付与の例



図 5 学習者のレビューの様子

改善点に対しては、2つの条件間で改善点の平均数に差が見られるかを、客観的データとして分析し、仮説3の検証に利用した。

4.4 実験結果

まず、仮説1の検証を行った。図6は、仮説1(エンゲージメント)に関するアンケート結果である。事後アンケートのQ1では、どちらの条件でも高い平均値を示していたが、有意差は見られなかった。全体アンケートのQ1でも、実験条件を選択した被験者が多かったが、両条件間に有意差は見られなかった。

次に、仮説2の検証を行った。図7は、仮説2(心理的抵抗感)に関するアンケート結果である。事後アン

ケートのQ9とQ10の平均値には、両条件で大きな差があり、対応のあるt検定を行ったところ、両条件の平均値に1%水準で有意差が見られた(Q9: $t(13)=3.47$, $p=0.004$; Q10: $t(13)=3.31$, $p=0.006$)。効果量を求めたところ、大きな値が得られた(Q9: $d=1.180$; Q10: $d=1.029$)。これらの結果は、ピアのロールでのセルフレビューが、セルフレビューにおける学習者の心理的抵抗感を軽減することを示唆しており、仮説2を支持している。

最後に、仮説3の検証を行った。図8と図9は、仮説3(客観視)に関するアンケート結果である。図10は、各条件における改善点の数の平均値である。事後アンケートでは、ほぼ全ての質問で、統制条件の平

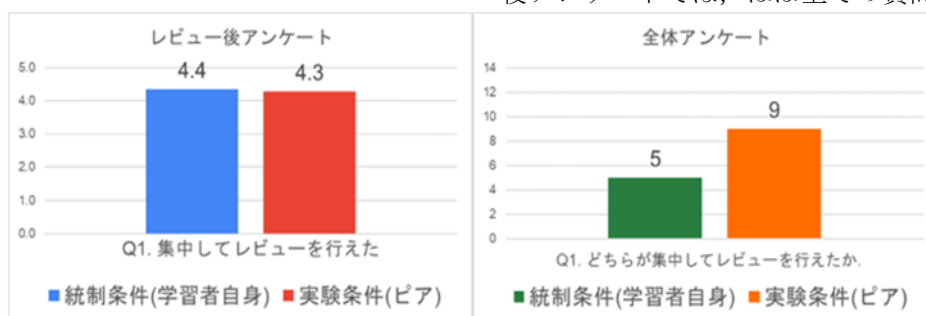


図6 エンゲージメント(仮説1)に関するアンケート結果

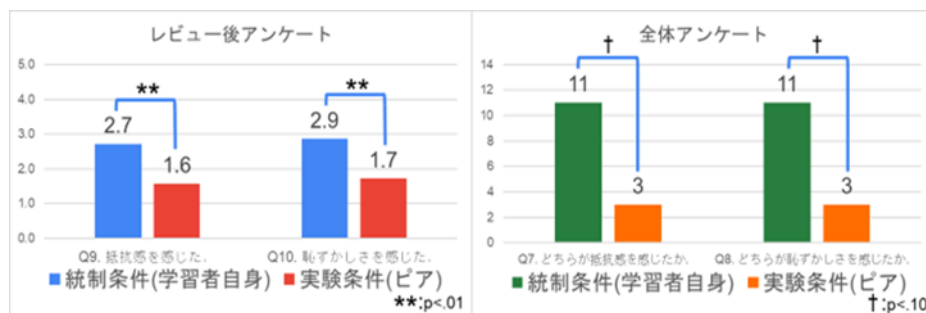


図7 心理的抵抗感(仮説2)に関するアンケート結果

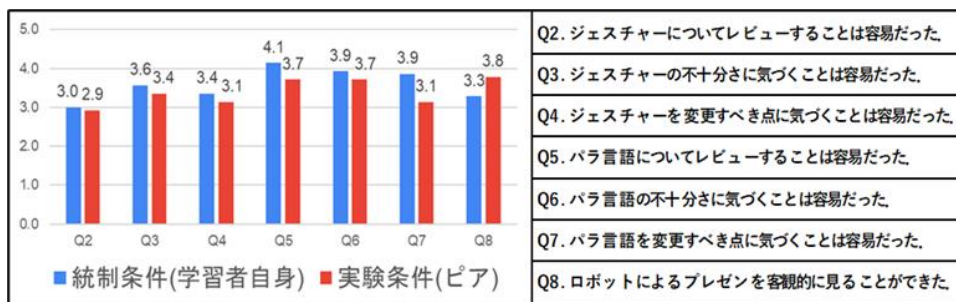


図8 客観視(仮説3)に関する事後アンケートの結果

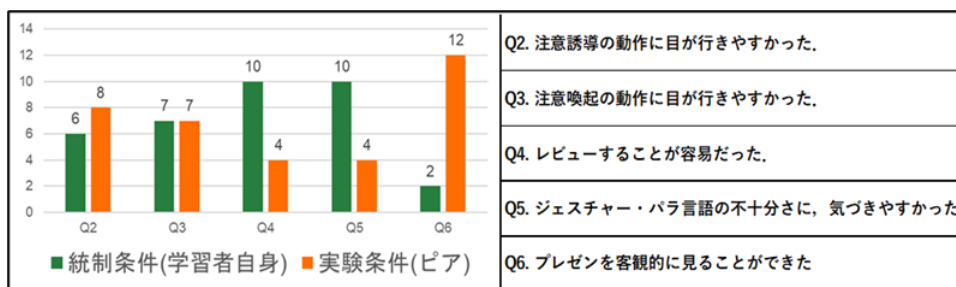


図9 客観視(仮説3)に関する全体アンケートの結果

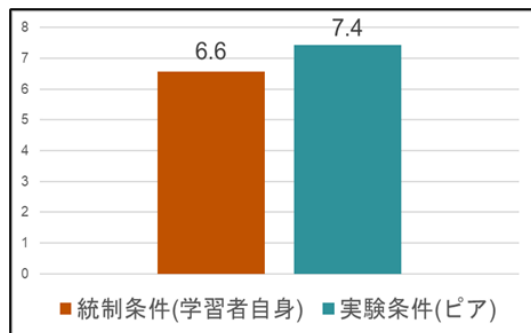


図 10 改善点の数の平均

均値が実験条件の平均値よりも大きくなっている。この結果は、統制条件の方が非言語動作の改善点を見つけやすいということを示しているが、有意差は見られなかった。一方、全体アンケートの Q6 では大きな差が見られ、14 人中 12 人の被験者が実験条件を選択した。この結果について、正確二項検定（両側検定）を実施した。その結果、5%水準で有意差が認められた（ $p=0.021$ ）。次に、図 10 に示す改善点数の分析を行った。両条件の平均値には差が見られ、実験条件の方が統制条件よりも大きな値となっていた。両条件の改善点の数について対応のある t 検定を行ったが有意差は見られなかった（ $t(13)=0.965$, $p=0.352$ ）。また、効果量も小さな値を示した（ $d=0.26$ ）。よって、これらの結果から仮説 3 が成り立つことは確認できなかった。

4.5 考察

事後アンケートと全体アンケートの結果から、ピアのロールでセルフレビューを行うことで、学習者の心理的抵抗感が軽減されることが示唆された。

エンゲージメントに関しては、アンケートの結果に有意差は見られなかったが、事後アンケートではどちらの条件でも高い平均値が見られた。統制条件でのセルフレビューが高いエンゲージメントを引き出すことは関連研究[3]でも示されている。そのため、本実験の結果からは、プレゼンテーションセルフレビュー支援ロボットのロールを変化させても、高いエンゲージメントが維持されることが示唆された。

最後に、仮説 3 の結果について考察する。今回の実験では、被験者は連続して 2 つのレビューを行った。2 つのプレゼンテーションは、ロボットのロール、声の高さ、色は異なるものの、同じ被験者のプレゼンテーションを再現したものであった。そのため被験者にとっては、1 回目よりも 2 回目の方がレビューしやす

く、改善点に気づきやすい状態であったと考えられる。このような順序効果は、図 11 に示す A 群の改善点の数において、実験条件（2 回目のレビュー）の平均値が統制条件（1 回目のレビュー）の平均値よりも高くなっていることから確認できる。この順序効果を考慮すると、B 群の改善点の数も 2 回目のレビューの方が、平均値が高いはずである。しかし、図 12 に示す B 群の改善点の数においては、実験条件（1 回目のレビュー）の平均値の方が統制条件（2 回目のレビュー）の平均値よりも高くなっている。これは、ピアのロールでのセルフレビューが、順序効果を打ち消すほど、改善点の気づきに貢献したためと考えられる。これらの結果は、ピアのロールでのセルフレビューが学習者の客観視を促進し、学習者の改善点への気づきを促したことを示唆していると考えられる。

5. 結論

ロボットを用いた学習支援研究として、筆者らは「講義ロボット」、「英文読み合いロボット」、「プレゼンテーションセルフレビュー支援ロボット」を開発してきた。これらの研究にてロボットは、学習者の教師、学習者のピア、学習者自身といったロールがあらかじめ与えられてデザインされており、これらのロールが学習者の学びに影響を与えている可能性がある。しかし、学習支援ロボットのロールによる学習者への影響につ

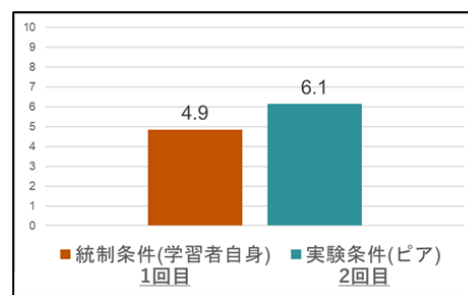


図 11 A 群における改善点の数の平均

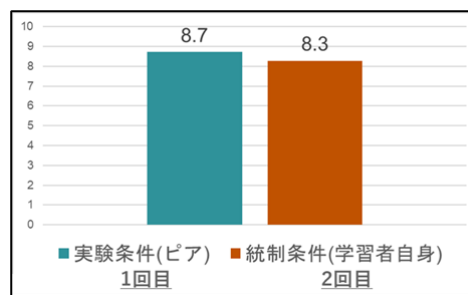


図 12 B 群における改善点の数の平均

いては、十分に検証されていない。

本研究では、学習支援ロボットのロールを変えたときの学習者の学びへの影響について検証を行った。「ロボットが学習者のプレゼンテーションを再現し、学習者自身がそれをレビューする」という文脈で、2種類のロール（「学習者自身」と「学習者のピア」）を用意し、検証実験を行った。ロボットが再現した同じ学習者のプレゼンテーションを「学習者自身」としてセルフレビューした場合と「学習者のピア」としてセルフレビューした場合を比較し、学習者のレビュー内容への影響を検証した。評価項目として、エンゲージメント、心理的抵抗感、客観視の変化に着目した。実験の結果、ピアのロールでのセルフレビューは、学習者の心理的抵抗感を軽減し、改善点への気づきを促進する可能性が示唆された。

今後、本研究で得られた結果を踏まえて学習者が自分のプレゼンテーションをピアのものとしてセルフレビューできるシステムを開発し、その評価を行いたい。また、筆者らが開発してきた「講義ロボット」や「英文読み合いロボット」など、ロボットを用いた他の学習支援においても、ロールの影響を確認することが必要である。

謝辞

本研究の一部は、JSPS 科研費 23H03505 の助成を受けた。

参 考 文 献

- (1) Ishino, T., Goto, M., Kashiara, A.: “Robot Lecture for Enhancing Presentation in Lecture”, Research and Practice in Technology Enhanced Learning, 17.1, pp.1-26. (2022)
- (2) 足立祥啓, 柏原昭博: 効果的な英文読み合いを行う学習パートナーロボットのデザインと評価, 教育システム情報学会第4回研究会, Vol.33, No.4, pp.85-92, 2018.
- (3) 柏原昭博, 稲澤佳祐: “プレゼンテーションロボットによるセルフレビュー支援”, 第82回先進的学習科学と工学研究会(SG-ALST), pp91-96(2018).
- (4) 加藤 由樹, 赤堀 侃司: 電子メールを使ったコミュニケーションにおける感情面に及ぼす相手の立場の影響, 日本教育工学会論文誌, 2006 年 29 巻 4 号 p. 543-557
- (5) Rod D. Roscoe & Michelene T. H. Chi (2008) : Tutor learning: the role of explaining and responding to questions, Instructional Science volume 36, p.321–350 (2008)
- (6) S. Kiesler, A. Powers, S. R. Fussell, and C. Torrey: Anthropomorphic Interactions with a Robot and Robot-like Agent, Social Cognition, Vol. 26, pp. 169-181 (2008)
- (7) A. Powers, A.D.I. Kramer, S. Lim, J. Kuo, Sau-lai Lee, S. Kiesler: Eliciting information from people with a gendered humanoid robot, ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005
- (8) Lee. SL, Lau. IYM, Kiesler. S, Chiu. CY: Human mental models of humanoid robots IEEE International Conference on Robotics and Automation Proceedings, 18-22 April 2005, p. 2767-2772
- (9) ERIC P. BETTINGER AND BRIDGET TERRY LONG: Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students, AMERICAN ECONOMIC REVIEW, VOL. 95, NO. 2, MAY 2005, pp. 152-157
- (10) Reich-Stiebert, N. Eyssel, F: (Ir)relevance of Gender?: On the Influence of Gender Stereotypes on Learning with a Robot, ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, pp. 166–176. ACM (2017).
- (11) Tatsuya Ishino, Mitsuhiro Goto, Akihiro Kashiara. “A Robot for Reconstruction Presentation Behavior in Lecture” 6th International Conference on Human-Agent Interaction(HAI), Southampton, UK (2018.12.16): pp.67-75.
- (12) 小林ひろみ, 小林めぐみ: “アカデミック・プレゼンテーション”, 株式会社朝倉書店, 東京 (2003)
- (13) 瀬谷遼太郎, 柏原昭博, “研究プレゼンテーションの診断に基づくロボットセルフレビュー支援”, 電子情報通信学会 教育工学研究会(ET), 信学技報, vol.119, No.236, ET2019-43, pp.63-68, 2019.
- (14) 恋声萌. 恋声,
<http://koigoemoe.g2.xrea.com/koigoe/koigoe.html>
(2023.6.7 確認).
- (15) VStone.: Sota, <https://sota.vstone.co.jp/home/>
(2023.6.6 確認).
- (16) Microsoft Corporation. 2018. Kinect for Xbox One | Xbox. <https://www.xbox.com/> (2023.6.7 確認)

インタラクティブロボット講義における ケーススタディ分析と講義インタラクショナルモデルデザイン

島崎俊介^{*1}, 柏原昭博^{*1}

^{*1} 電気通信大学 大学院情報理工学研究科 情報学専攻

Evaluating Case Studies and Designing Interaction Model for Interactive Robot Lecture

Toshiyuki Shimazaki^{*1}, Akihiro Kashiara^{*1}

^{*1} Graduate School of Informatics and Engineering, The University of Electro-Communications

筆者らはインタラクティブロボット講義研究において、理工系大学生や高校生を被験者としたケーススタディを実施し、学習状態推定の精度向上、ロボットによる講義スライドとインタラクションの適応的制御を通じて、学習者の講義に対する注意制御、注意維持及び理解促進の支援を試みた。本稿では、インタラクティブロボット講義のケーススタディを総括することで、学習者のエンゲージメントを促進し、長時間の講義受講時の注意維持・理解促進を支援する際の課題を整理する。また、学習者のエンゲージメントをより高めるための、学習者からロボットへの働きかけを含むモデルのデザインも言及する。

キーワード: インタラクティブロボット講義, 学習状態推定, 注意・理解リカバリ,
インタラクションデザイン

1. はじめに

人間とロボットのインタラクションを研究するヒューマンエージェントインタラクション(HAI)分野や、ChatGPTに代表される大規模言語モデルによるAI対話システムの急速な発展により、人間とロボットが共生するロボット社会の実現が現実的となっている。ロボットの社会実装として、二足歩行可能で特定のスポーツができるロボットや、人間のような表情を有するロボットが開発されている。近い将来には、携帯電話、PC、タブレット端末が普及したのと同じように1人1台パートナーロボットの普及も現実的と考えられる。

筆者らは、将来的な1人1台学習パートナーロボット時代の到来を目指した学習支援ロボット研究の1つとして、人間講師と同等の講義をロボットが代行できないか？という学術的な問いを掲げ、ロボットによる講義に関する研究を進めている。講義では、通常長時間に及ぶほど、講義に対する学習者の注意を制御する

こと（以下、注意制御）や、理解状態を把握しながら、インタラクティブに進めることが重要となる。しかしながら、熟達した人間講師でもそれらを長時間遂行し続けることは負荷があり、容易ではない。そこで筆者らは、先行研究でロボット講義を実現した^{1,2)}。

ロボット講義とは、学習支援ロボットが人間講師の不十分な講義動作（指差し、ジェスチャー、パラ言語の非言語動作）をモデルベースに再構成し、人間講師の講義代行を可能とした研究である。成果として、短時間の講義に対する学習者の注意制御と理解促進の効果を実証した²⁾。しかし、学習者の受講状態を考慮していないため、講義が長くなるほど学習者の注意が外れ、講義に追従できず理解できない可能性がある。

筆者らは、上記問題点の解決として、講義受講中に計測された学習者のセンシングデータに基づく学習状態推定と、受講状態に応じたロボットによる適応的なインタラクションを行うことで、長時間の講義における学習者の注意維持と理解追従を可能にするインタラ

クティブロボット講義研究とシステム開発を進めている³⁻⁹⁾。本システムは、人間講師の講義の再現性を高めるため、SoftBank Robotics 社製のヒューマノイドロボットの NAO を用いており、学習者に近づくウォークが可能で、学習者の注意維持支援に効果的であることを確認している^{4,5,6)}。同時に、Kinect による学習者の受講状態を集中状態・非集中状態の 2 状態で推定していたが、集中状態であっても学習者の理解状況の把握が困難であった。そこで、学習者の注意回復（以下、注意リカバリ）と理解回復（以下、理解リカバリ）を行うアルゴリズムと、講義内容の重要箇所や関係を問うチャットボット機能を具備した新たなインタラクティブロボット講義システムを開発した⁷⁾。ケーススタディでは、学習者の学習状態推定のセンシングに基づく注意リカバリの効果を確認している^{8,9)}。

本稿では、筆者らが実施してきたインタラクティブロボット講義におけるケーススタディを分析し、インタラクティブロボット講義の課題を整理する。また、学習者の講義に対する積極的関与(エンゲージメント)をより高めるため、学習者からロボットへの働きかけを含むインタラクションのモデルについて言及する。

以下、2 章ではインタラクティブロボット講義の位置づけとモデル、アルゴリズム、システムの枠組みを述べる。3 章では、今まで実施したケーススタディの結果と考察を示し、4 章ではケーススタディの結果を総括し、インタラクションモデルの新たなデザインについて述べる。5 章では、今後の課題を述べる。

2. インタラクティブロボット講義

2.1 本研究の位置づけ

本研究では、人間講師の個性を活かし、講義動作やスタイルを反映することを意図して、ロボットが講義を代行することを目指している。そのため、講師の講義を収録して、ロボットが再現するとともに不十分・不適切な講義動作をモデルベースに再構成する。そして、講義中の学習者の受講状態をリアルタイムで推定し、必要なタイミングでロボットからインタラクションを行うことで学習者の注意・理解リカバリを行う。

また、現時点ではスライドベースの講義を前提に、ヒューマノイドロボット 1 台に対して受講生 1 名を想

定しており、先行研究のモデルとアルゴリズム^{5,6,7,8)}に基づきインタラクティブな講義を実現している。

ロボット講義の関連研究の多くは、ベースの人間講師が存在せず、ロボット講師を新たに作成する研究が多い。例えば、文献¹⁰⁾では、ソーシャルロボットの Sota を用いて認知的徒弟制理論に基づく教師型ロボットを開発し、従来型の解き方を教授する教師型ロボットに比べ、中学生を対象とした数学の学習効果の向上を報告しているが、特定の人間講師の講義代行ではない。他の関連研究においても、特定の学習者に限定した教師型ロボット¹¹⁾、や特定の講義内容に特化したロボットを開発¹²⁾しているが、本研究では、対象の学習者・講義内容を限定していない。これにより、モデルとなる人間講師が講義しているような存在感や、非言語動作による学習効果の向上を期待している。また、講義では学習支援メディアとしてのロボットの利点である擬人化傾向と身体性¹³⁾を最大限に活かすことができると考えている。そのため本研究では、ヒューマノイドロボット NAO を用いている。

また現時点で、ロボット講義の関連研究において学習者の受講状態をセンシングし、学習状態推定を試みている研究はない。本研究では、学習状態推定に応じて適応的に講義スライドのシーケンスを制御し、学習者に適応的なインタラクションを行い長時間の講義における注意維持・理解追従を実現しようとする新規性の高い研究といえる。

筆者らが開発しているインタラクティブロボット講義システムでは、まず人間講師が作成した講義シナリオをもとに人間講師の講義動作を診断・再現・再構成するためのモデル（講義動作モデル）を用いて、ロボットが人間講師の講義代行を可能にする。次に、学習者の姿勢・視線を受講状態のデータとしてセンシングし（姿勢・視線センシングデータ）、学習状態推定を行う。その上で、学習者の受講状態に応じてロボットが講義シナリオの順序（シーケンス）を制御するためのモデル（シナリオ制御モデル）と、ロボットが学習者の注意及び理解をシステムティックにリカバリするためのアルゴリズム（注意・理解リカバリアルゴリズム）を組み合わせ、学習者とのインタラクションを実施する。以下、システムの詳細について述べる。

2.2 講義シナリオ

本研究における講義シナリオとは、複数の切れ目(セグメント)からなる講義スライドの系列、人間講師による講義意図、講義スライド毎の非言語動作・パラ言語・スライド内問題、セグメント内のスライド間問題から構成されている。講義意図は、人間講師が講義を行う意図を表し、講義全体で1つに固定されるものとする。スライド内問題とは、特定のスライドにおける重要箇所を問う問題であり、スライド間問題とは、複数のスライドの関係性を問う問題である。これらの問題は、講義シナリオ作成時に予め人間講師が作成することとし、各講義スライドで重要なキーワードを問うことで学習者の理解状態を把握して、理解リカバリにおける問題提示に活用することを想定している。

2.3 講義動作モデル

講師が講義を意図通りに実施するには、講義を受講する学習者の状態を仮定し、講義中の指差し、ジェスチャー、パラ言語といった講義の非言語動作(講義動作)を組み立てる必要がある。筆者らは、これまでロボットが講義中に適切な講義動作により学習者の受講状態を集中状態に意図的に遷移させるための講義動作モデル(図1)をデザインしてきた。本モデルによって、事前に収録された人間講師の講義を診断し、不十分な講義動作について講義意図から適切な非言語動作を導き再構成することができる。本システムでは、このようにモデルベースに人間講師による講義代行を実現する。

2.4 シナリオ制御モデル

ロボット講義では、ロボットが学習者の受講状態に応じて講義スライドの説明順序を変更し、学習者の注意・理解リカバリするためのインタラクションを行う。特に、長時間の講義ほど、学習者に応じた講義シナリオの適応的な制御が必須である。そこで筆者らは、学習者の状態に応じた注意・理解リカバリを実現するためのシナリオ制御モデル(図2)をデザインした⁹⁾。本モデルは、学習者の受講状態をトリガーにシーケンスを制御することで、学習者の講義に対する注意喚起を行う。本モデルは、学習者の受講状態、講師の講義意図の変化、講義スライドシーケンス制御、注意リカバリ動作カテゴリ、インタラクションのための非言語動作

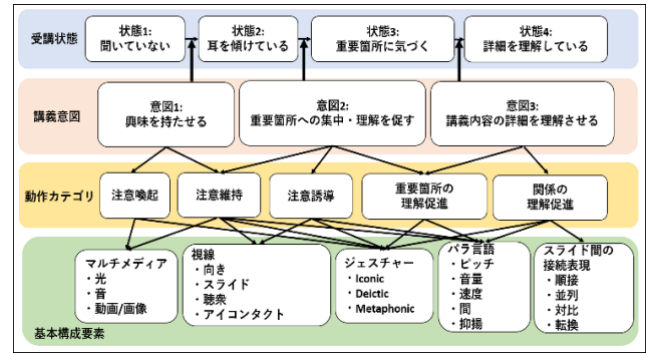


図1 講義動作モデル

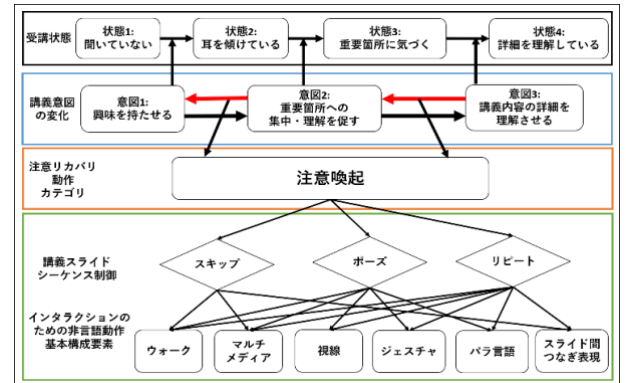


図2 シナリオ制御モデル

基本構成要素の5層から構成される。受講状態は、講義動作モデルと同様に4状態で、状態1を非集中状態、状態2以上を集中状態とし、姿勢・視線センシングデータに基づく学習状態推定を行う。

2.5 注意・理解リカバリアルゴリズム

インタラクティブロボット講義における注意・理解リカバリは、図3に基づき実施される。本研究では現在のところ、学習者の受講状態を状態2: (耳を傾けている)、講義意図を意図2: (重要箇所への理解・集中を促す)として状態3へ移行させる講義を想定している。

注意リカバリは、学習者の受講状態が状態1: (聞いていない) 場合をトリガーとして起動される。図3の太枠の通り、シナリオ制御モデルに基づき講義スライドの説明を中断し、例えばウォークで学習者に近づくことで講義への注意喚起を行う。その後、ロボットは学習者に講義スライドをもう一度説明するか問いかけ、学習者は説明要否を選択できる。学習者が説明を要とした場合、ウォークとマルチメディアによる注意喚起実施後、ロボットは1回限定で、誇張して講義スライドの説明をリピートする。そして、受講状態が初期状態の状態2に回復したとみなし、意図2での講義を再開して、注意リカバリが完了する。なお注意リカ

バリから理解リカバリへ遷移する条件は、予め人間講師が講義シナリオ作成時に意図した、同一セグメント内での注意リカバリの回数 n をタイミングとし、チャットボット機能を起動するトリガーとなる。

理解リカバリは、受講状態が状態 1 ではない、かつ注意リカバリの回数 n 、または講義スライドのセグメントをトリガーとして、図 3 の点線枠の通り、チャットボット機能が起動される。チャットボット機能によるテスト問題は、4 択の選択式問題で与えられ、スライド内問題とスライド間問題を用意する。本研究では、スライド内問題を全問正解すると状態 3: (重要箇所に基づく) と推定し、またスライド間問題を全問正解すると状態 4: (詳細を理解する) と推定する。

2.6 システムの枠組み

図 4 にインタラクティブロボット講義システムの枠組みを示す。本システムは、講義シナリオを生成する講義収録サブシステム、学習者の姿勢・視線センシングデータの計測と学習状態推定を行う学習状態センシングサブシステム、学習状態推定に基づき講義シナリオ制御、注意・理解リカバリインタラクションを実施する動的シナリオ・インタラクション生成サブシステム、ロボットが講義代行・インタラクションを行う講義ロボットサブシステムの 4 つから構成されている⁹⁾。

講義収録サブシステムは、人間講師が行う講義を収録する。収録は、講義スライド、非言語動作、パラ言語、スライド内問題、スライド間問題を収集し、非言語動作が不十分な点は、講義動作モデルに基づき診断・再構成を行うことで講義シナリオを生成する。

学習状態センシングサブシステムは、姿勢センシングと視線センシングの並列処理を行う。姿勢センシングでは、Kinect を用いて学習者の骨格や顔のトラッキングデータを画像形式で取得する。その後 10 秒毎のサンプリング間隔で姿勢推定フレームワーク OpenPose の信頼度を取得する。信頼度は、右耳、左耳、右目、左目、右手、左手、首を使用する。本研究では、講義スライド毎に姿勢センシングを行い、学習者の顔が右向き、左向き、上向き、下向きのいずれかで、同じ講義スライド内で 2 回以上推定された学習状態を非集中状態と推定し、ロボットが学習者の注意が外れたタイミングと判断する。例えば、顔の右向きは

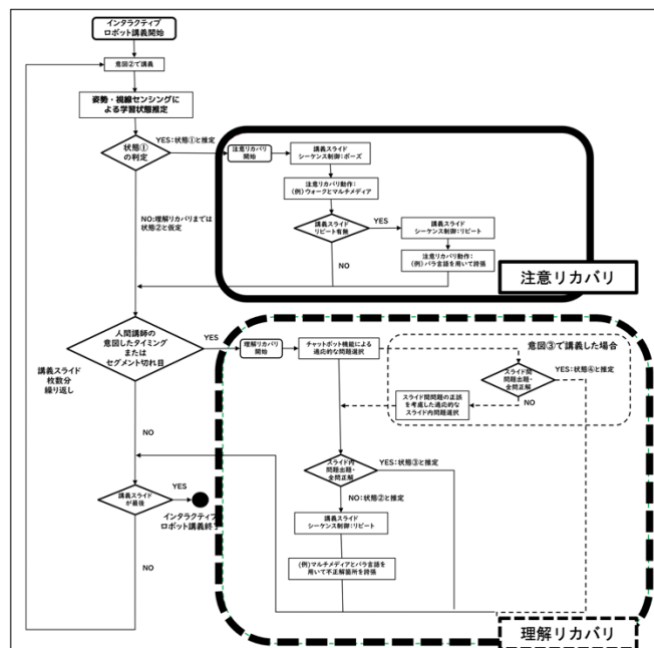


図 3 注意・理解リカバリアルゴリズム

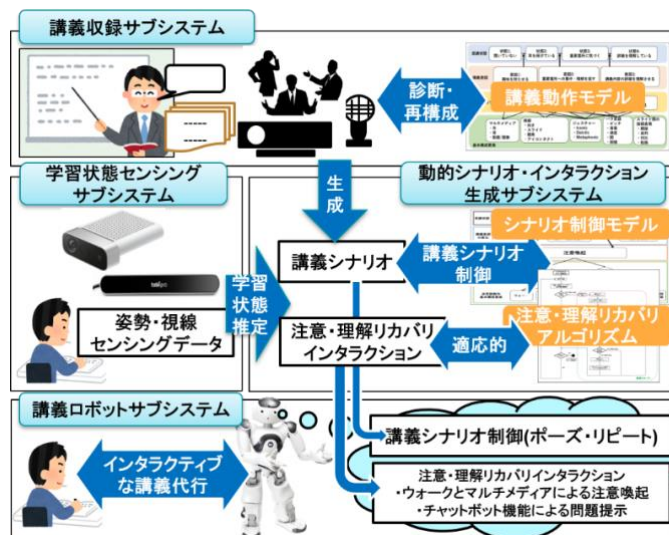


図 4 インタラクティブロボット講義システム

受講中に取得した学習者の右耳と右目の両方の信頼度の値が、講義前のキャリブレーションで取得した信頼度の平均値より低い場合に推定される。本手法は、多様な受講時の姿勢に頑健であることを確認している⁹⁾。視線センシングでは、tobii pro nano を用いて講義における学習者の PC モニター閲覧状態を取得し、講義動作モデル及びシナリオ制御モデルの学習者の受講状態が、状態 1(非集中状態)あるいは状態 2 以上(集中状態)のいずれかを推定する際に用いる。

本手法は、視線情報を用いた関連研究の学習状態推定と異なる点として、あくまで学習者が PC モニターを見ているか・否かの 2 値に限定している。

動的シナリオ・インタラクション生成サブシステムは、学習者の受講状態が想定された状態と異なる場合、

講義スライドシーケンス制御を行い、注意リカバリのためのインタラクションと、理解リカバリのためのチャットボット機能による適応的問題提示を行う。チャットボットが提示するテスト問題そのものは、予め人間講師が用意するが、適応的問題提示では、注意リカバリの回数 n に応じて、注意リカバリが発生した講義スライドに応じて学習者毎にテスト問題を出题する。

講義ロボットサブシステムは、講義シナリオをベースに NAO が人間講師の講義を代行し、受講状態に応じた注意・理解リカバリのインタラクションを行う。

3. ケーススタディ

筆者らはこれまで理工系大学生及び高校生を対象としたケーススタディを実施してきた¹⁻⁹⁾。本稿では、未報告の高校生を対象としたケーススタディを述べる。

3.1 目的と仮説

インタラクティブロボット講義が、注意・理解リカバリ支援に有効かどうかを検証するために、以下の仮説を検証するケーススタディを実施した。

H1: 講義スライドのシーケンス制御による注意喚起のインタラクションは、長時間の講義における注意リカバリに有効である。

H2: チャットボット機能による問題提示・出題は、長時間の講義における理解リカバリに有効である。

H1 を検討する評価指標として、注意リカバリ・理解リカバリの回数、受講後アンケート、全体アンケートによる主観的評価を行った。

H2 を検証する評価指標として、受講後の理解度テストを用いた客観的評価と、受講後アンケート、全体アンケートによる主観的評価を行った。

3.2 実験方法

実験被験者は、公立高校生 16 名であり、講義は情報科の内容を題材とし、高校教員による講義を予め収録した 30 分程度の 2 つのコンテンツとして情報社会における法(以下、情報社会)とネットワークのしくみ(以下、ネットワーク)を用意した。実験では、被験者を 8 名ずつの二群に分け、被験者内計画で実施した。その際、インタラクティブに注意・理解リカバリを行うインタラクティブロボット講義条件(IL 条件・実験条件)

と、ロボット講義条件(NL 条件・統制条件)の 2 条件を設定し、受講順序のカウンターバランスをとった。被験者には、任意で講義中の講義メモによるノートテイキングを許可した。実験中の講義動画は、被験者と高校教員に同意を取り、受講状態分析の目的で撮影した。なお、両条件とも、姿勢・視線センシングデータを取得し、IL 条件での注意・理解リカバリに用いた。なお、実験実施にあたっては、電気通信大学人を対象とする研究に関する倫理委員会の承認を受けた。

受講後は、講義メモを回収した。また、被験者は講義毎に、スライド内問題 10 問の理解度テスト(選択式問題 9 問 1 点・記述式問題 1 問 2 点の計 11 点満点、回答時間は 10 分)、7 段階リッカー尺度による受講後アンケートに回答し、最後に二者択一による全体アンケートに回答した。理解度テストでは、チャットボット機能と異なる問題を提示した。受講後・全体アンケート項目は、HRI 分野の主観評価の心理尺度である GodSpeed Questionnaire を参考とし、ロボットによる講義に関する設問(Q1~6)、ロボットによるインタラクションに関する設問(Q7~9)、ロボットによる擬人性に関する設問(Q10-14)、講義全体に関する設問(受講後アンケート: Q15-17, 全体アンケート: Q15-16)からなる。ケーススタディでは、図 3 の学習者の受講状態を状態 2、講義意図を意図 2 とした。チャットボット機能のトリガーは、注意リカバリの回数 $n=2$ とした。

3.3 結果

図 5・6 に各講義に対する理解度テスト点数、図 7・8 に講義受講後のアンケート平均値、図 9 に全体アンケート結果を示す。実験では、IL 条件で注意リカバリが 6 回、理解リカバリが 2 回確認され、NL 条件で注意リカバリが 1 回、理解リカバリが 0 回確認された。

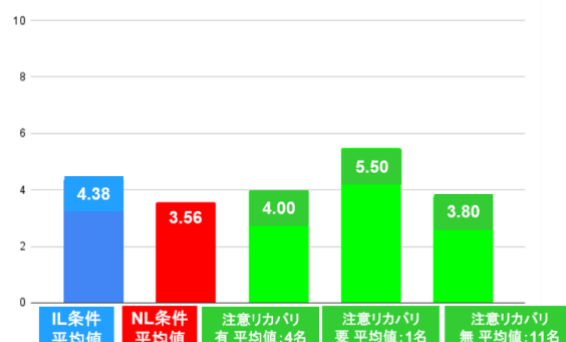


図 5 理解度テストの平均点 (情報社会)

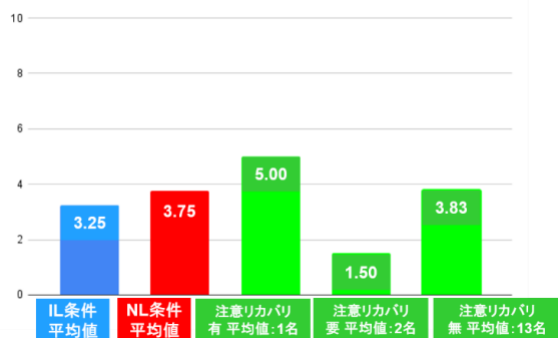


図 6 理解度テストの平均点（ネットワーク）

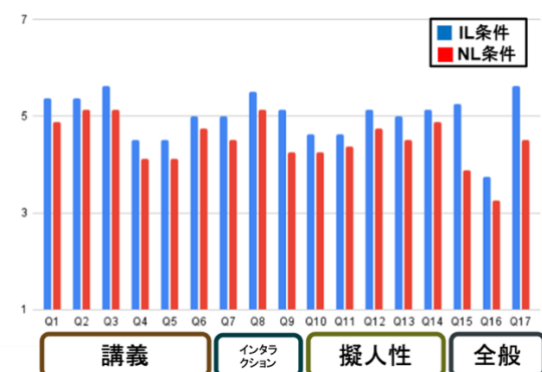


図 7 受講後アンケートの平均値（情報社会）

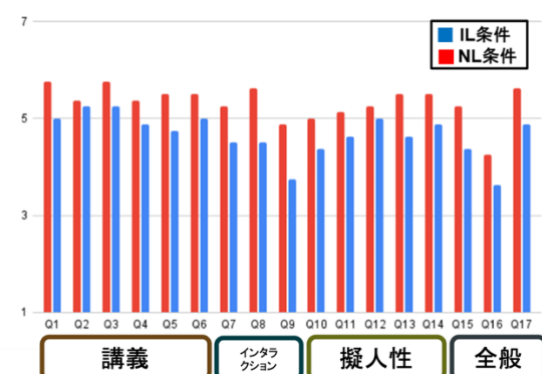


図 8 受講後アンケートの平均値（ネットワーク）

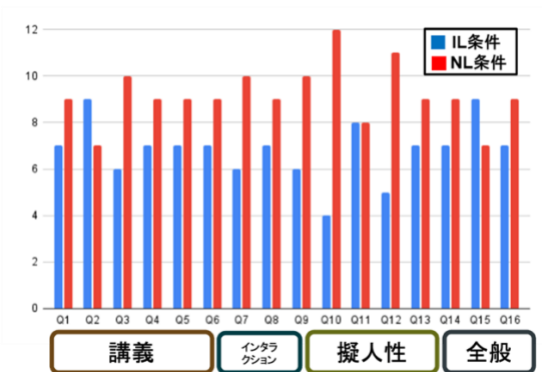


図 9 全体アンケートの結果

3.4 考察

まず、図 5, 6 より情報社会では、IL 条件の平均値が 4.38 で、NL 条件の平均値が 3.56 であった。有意差は見られなかったものの、IL 条件の値が高い結果と

なった。一方、ネットワークでは、情報社会では、IL 条件の平均値が 3.25 で、NL 条件の平均値が 3.75 で NL 条件の値が高かった。

次に、情報社会とネットワークの理解度テストを注意リカバリが発生した被験者群（注意リカバリ有群）、NL 条件でシステムは注意リカバリを必要と判断できた被験者群（注意リカバリ要群）、注意リカバリが発生しなかった被験者群（注意リカバリ無）の 3 群にわけ、理解度テストの点数を分析した。

情報社会では、図 5 より、被験者 16 名中、注意リカバリ有群が 4 名、注意リカバリ要群が 1 名、注意リカバリ無群が 11 名だった。情報社会の注意リカバリ有群の平均点は 4.00 で、有意差は確認できなかったものの、注意リカバリ無群の平均点 3.80 より高かった。この結果は、注意リカバリのインタラクションの効果が示唆される。また、理解リカバリ有群で理解リカバリが発生した 2 名の平均点はそれぞれ、4.5 点と 3.5 点であり、IL 条件の平均値に近い結果からも、注意リカバリに加えて、理解リカバリによる問題提示の効果が示唆される。また、注意リカバリ要群 1 名に着目すると、平均点 5.50 で他の 2 群よりは高かったが、講義動画を確認すると全体的に集中して講義を聞いていないように見受けられた。この被験者に注意リカバリが発生すると更に理解度テストの点数が高くなった可能性も考えられる。

一方、ネットワークでは、図 6 より、被験者 16 名中、注意リカバリ有群が 1 名、注意リカバリ要群が 2 名、注意リカバリ無群が 13 名だった。ネットワークの方で注意リカバリ発生数が少ない理由として、IL 条件と NL 条件の両方で、ネットワークは 2 つめに受講するコンテンツだったため、受講前に休憩があったこと、情報社会のコンテンツで注意リカバリが最初に発生することが注意維持の要因になったと考えられる。また、注意リカバリ有群の 1 名の点数は、注意リカバリ無群の 13 名の平均点より高いため、注意リカバリのインタラクションの効果が示唆される。次に、注意リカバリ要群の 2 名の点数に着目すると、それぞれ 3 点と 0 点であったため、図 6 の平均点が 1.5 点となっていた。この 0 点の被験者は、情報社会(IL 条件)で理解リカバリが発動して 3.5 点になった被験者であった。0 点だった理由として、ネットワーク(NL 条件)では注意・理

解リカバリが働かなかったため、講義の注意維持・理解追従ができなかったと考えられるが、唯一の0点の被験者であるため、他の15名と比較して外れ値であるとも考えられる。講義動画を確認しても、全体的に上の空状態で受講していた。

情報社会とネットワークの両方で0点の被験者を除く15名に着目すると、注意リカバリが1回発生した7名中6名は理解リカバリが発生せず、理解リカバリが発生した1名は、平均点を上回っていた。これらより、注意リカバリが1回発生するだけでも講義に対する注意維持に効果が示唆され、理解リカバリが発生すると、理解促進の効果が示唆されると考えられる。ただし、仮説H1とH2の検証は、今後もケーススタディを積み重ねて実施する必要がある。

次に、図7・図8の受講後アンケートと図9の全体アンケートの考察として、条件間に有意差は見られず、ネットワークと全体アンケート項目ではNL条件のほうがIL条件より全体的に高い値となった。理由として、注意リカバリが発生しなかった受講状態が良好な被験者にとっては、IL条件とNL条件の両方でロボットによるインタラクションが発生せず、条件間の差が不明だったと考える。今後はIL条件においては、講義スライドのセグメントの切れ目で、ロボットが学習者にインタラクションを行い、講義スライドのリピート・ポーズ、チャットボット機能による問題提示の可否を示し、条件間の差異を明確にする必要がある。

最後に、推定精度の評価として注意リカバリが発生した被験者7名に着目し、高校教員による推定結果を正解とし、システムによる推定結果の適合率・再現率を注意リカバリが発生されたコンテンツの講義スライド全体の値で算出した。結果は、人間講師の推定精度と比較し、概ね良好であった。詳細は文献¹¹⁾に譲る。

4. エンゲージメント促進のための講義インタラクションモデル

本章では、これまでのケーススタディを総括し、インタラクティブロボット講義の課題を整理し、学習者の講義におけるエンゲージメントをより促進するためのインタラクションモデルデザインについて述べる。

4.1 インタラクティブロボット講義の課題

これまでのケーススタディの結果^{2~10)}を総括すると、次の2点が今後解決すべき研究課題であるといえる。

1) ベースとなるロボット講義(インタラクションを含まない)の注意制御効果が高く、注意リカバリが発生しにくい、

2) 実験室実験だけでは注意・理解の評価が困難

1)については、実験条件をインタラクティブロボット講義、統制条件をロボット講義としてケーススタディを実施してきたが、現状ではロボット講義条件においても人間講師の非言語動作の再構成は共通なため、講義に対する注意制御が効いていると考えられる。

2)については、実験室実験では受講中の姿勢が普段の講義より正しくなる傾向にあり、姿勢センシングだけでは非集中状態の推定は困難と考えられる。検討が必要なのはノートテイキングである。ケーススタディでは、システムは顔が下向きのノートテイキングを非集中状態と推定したが、高校教員は集中状態とした。

また、現状のインタラクティブロボット講義では、学習状態推定に応じたロボットから学習者への適応的な働きかけは実現できているものの、学習者からロボットへの働きかけは想定外となっている。その結果、講義中の姿勢・視線センシングデータでは受講状態が良好と推定された場合でも、学習者は注意が外れている場合があり、注意維持・理解追従の支援が十分ではない。そこで、インタラクティブロボット講義においても、学習者が講義に非集中状態と感じた場合にロボットに働きかけられる枠組みをモデルに組み込むことを検討する。これにより、より学習者のエンゲージメント促進が期待でき、先の課題解決も期待できる。

4.2 講義インタラクションモデル

以上の議論に基づきデザインした現時点の講義インタラクションモデルを示す(図10)。本モデルは、学習者の受講状態、学習者のインタラクション意図、学習者のインタラクション動作の3層から構成されている。

学習者の受講状態は、非集中状態(講義動作モデル・シナリオ制御モデルの状態1)と集中状態(講義動作モデル・シナリオ制御モデルの状態2,状態3,状態4)の2値とした。理由として、学習者が受講中に学習状態を4状態から同定するのは困難と想定したからである。

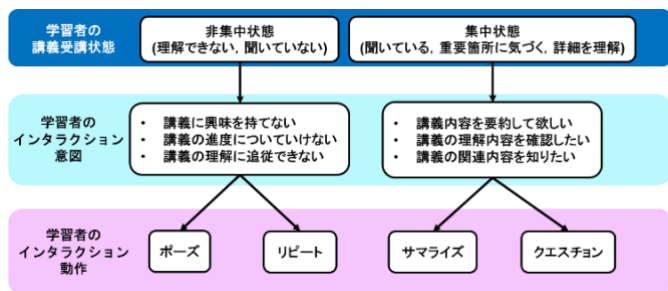


図10 講義インタラクションモデル

学習者のインタラクション意図では、学習者がどのような意図でロボットにインタラクションするかを整理している。具体的には、非集中状態では、学習者が講義に興味を持ってない、講義の進捗についていけない、講義の理解に追従できない、集中状態では、講義内容を要約して欲しい、講義の理解内容を確認したい、講義の関連内容を知りたいとした。

学習者のインタラクション動作では、具体的に学習者がロボットに対して行うインタラクション動作を記述している。学習者が非集中状態と感じた時は、ロボットに対して、シナリオ制御モデルと同一のポーズとリピートのインタラクションを想定する。学習者が集中状態と感じた時は、ロボットに対して、今まで説明した講義スライドの内容をまとめるサマライズと、今まで説明した講義スライドに関して質問を行うクエスチョンの2種類を想定する。これらは、技術的に実現することが容易でないが、人間講師がある程度、講義スライドの要約や、想定される質問などを事前に用意することと、生成AIが回答可能な内容に限定することで実現できると考える。

5. おわりに

本稿では、インタラクティブロボット講義研究のケーススタディ分析を行い、講義インタラクションモデルデザインを述べた。今後も引き続き、モデルの洗練とシステムの改良を行い、ケーススタディを実施してインタラクティブロボット講義の有用性を評価する。

謝辞

本研究の一部は、JSPS 科研費 23H03505 の助成を受けた。

参考文献

- (1) 柏原昭博, 島崎俊介: “ロボット講義におけるインタラクションデザイン”, 教育システム情報学会2022年度第2回研究会, pp.16-20, (2022)
- (2) Tatsuya Ishino, Mitsuhiro Goto, and Akihiro Kashiwara: “Robot Lecture for Enhancing Presentation in Lecture” Journal of Research and Practice in Technology Enhanced Learning, pp.1-22, (2022)
- (3) 柏原昭博, 菅原歩夢: “ロボット講義における講義シナリオの適応的制御”, 教育システム情報学会 2020 年度第6 回研究会, pp.19-26 (2021)
- (4) Shimazaki, T., Sugawara, A., Goto, M., and Kashiwara: “An Interactive Robot Lecture System Embedding Lecture Behavior Model”, Lecture Notes in Computer Science, vol 13329. Springer, (2022)
- (5) 島崎俊介, 菅原歩夢, 後藤充裕, 柏原昭博: “ヒューマノイドロボットを用いたインタラクティブロボット講義システム”, 電子情報通信学会研究会, pp.97-102 (2022)
- (6) 島崎俊介, 菅原歩夢, 後藤充裕, 柏原昭博: “インタラクティブロボット講義における学習状態推定と注意維持支援”, 教育システム情報学会全国大会, pp.155-156, (2022)
- (7) 島崎俊介, 柏原昭博: “インタラクティブロボット講義における注意・理解リカバリとその評価”, 教育システム情報学会2022年度第4回研究会, pp.70-77, (2022)
- (8) 島崎俊介, 柏原昭博: “インタラクティブロボット講義における注意・理解リカバリの有効性評価”, 教育システム情報学会2022年度第6回研究会, pp.149-156, (2023)
- (9) 島崎俊介, 柏原昭博: “学習者の姿勢・視線センシングデータを用いた状態推定によるインタラクティブロボット講義の評価”, 教育システム情報学会全国大会, (2023)
- (10) 宮内建弥, ジメネスフェリックス, 吉川大弘, 古橋武, 加納政芳: “認知的徒弟制理論に基づき教示を行うロボットと中学生における共同学習”, 知能と情報(日本知能情報ファジィ学会誌). 31 巻 5 号 pp. 834-841, (2019)
- (11) 前出晃佑, 増田寛之, 布施陽太郎, 高木昇, 澤井圭, 本吉達郎: “ロボット講義において参加者の自発的なコメントを促すためのシステム開発”, 第38回ファジィシステムシンポジウム. pp. 339-343, (2022)
- (12) 島谷二郎, Palinko Oskar, 吉川雄一郎, 陣内寛大, 小川浩平, 石黒浩: “ロボット質問支援システムによる発言促進: 高校生・大学生を対象とした講義における実証実験”, ヒューマンインタフェース学会論文誌, pp.369-380, (2020)

心拍変動の計測によるビジュアル型言語とテキスト型言語の 学習時の精神的ストレス分析について

梅澤 克之^{*1} 越川 拓海^{*2} 中澤 真^{*3} 平澤 茂一^{*4}

^{*1} 湘南工科大学 ^{*2} 旭情報サービス（株） ^{*3} 会津大学短期大学部 ^{*4} 早稲田大学

About the Mental Stress Analysis during Learning of Visual- and Text-based Programming Language by Measuring Heart Rate Variability

Katsuyuki Umezawa^{*1} Takumi Koshikawa^{*2} Makoto Nakazawa^{*3}
Shigeichi Hirasawa^{*4}

^{*1} Shonan Institute of Technology ^{*2} Asahi Intelligence Service Co., Ltd.
^{*3} Junior College of Aizu ^{*4} Waseda University

Our research project aims to establish a methodology for the transition between the two types of languages. In our previous research, we proposed an intermediate content with intermediate features between visual-based language and text-based language. And we showed that the comprehension of the text-based programming language is improved by inserting the intermediate content learning between the learning of both types of languages. In this study, we measure the heart rate variability of the learner during the learning of the visual-based programming language and the text-based programming language, and confirm whether there is any difference in the mental stress during the learning of both types of programming languages. Experiments confirmed that the heart rate variability value decreased (stress level increased) during visual-based language learning, especially in subjects accustomed to keyboard input.

キーワード：ビジュアル型言語，テキスト型言語，学習分析，心拍変動，精神的ストレス

1 はじめに

近年，プログラミングの入門としてビジュアル型のプログラミング言語（以降，ビジュアル型言語と呼ぶ）が使われるようになってきている．その後はC言語やJava言語などのテキスト型プログラミング

言語（以降，テキスト型言語と呼ぶ）に移行していくことになる．しかしシームレスな移行方法は確立されていない．

我々は，ビジュアル型言語からテキスト型言語への移行の方法論を確立することを目的とする研究プロジェクトを開始している．具体的には，ビジュア

ル型言語とテキスト型言語の学習の利点を有し、両者の差異を埋める教育コンテンツ（中間型言語とよぶ）を検討・試作し、実証実験を通して、学習結果だけでなく学習中の学習状態を評価する。これにより、その中間型言語の有効性を評価し、今後の初等・中等プログラミング教育に役立つ教育コンテンツ（中間型言語）を完成させる。ビジュアル型言語とテキスト型言語の溝を埋める想定の中間型言語の評価に関しては、今までは、学習後のアンケートや成績など、学習効果の結果のみに着目しており、理解できるようになったか否かを評価するのみであったが、これらの評価方法では、今回提案する中間型言語の効果を正確に測ることが出来ない。我々の研究プロジェクトでは、学習後の従来の評価方法に加えて、学習中の脳波や視線、表情などの生体情報および学習中の学習履歴を用いて学習状態を計測し、中間型言語がビジュアル型言語とテキスト型言語の中間的な役割を果たし、スムーズな移行に貢献しているかを分析・評価する。本研究が確立されると、プログラミング言語の初学者がビジュアル型言語による学習から始めて、その後、シームレスかつ自発的にテキスト型言語の学習に遷移できるようになることが期待できる。

我々は以前の研究⁽¹⁾で、上述の研究プロジェクトの一部として、まず中間コンテンツの提案に着目し、ビジュアル型言語の学習とテキスト型言語の学習の間に、我々が提案する中間コンテンツを用いた学習を挟む方がその後のテキスト型言語の理解度が高まるということを実証実験を通して示した。さらに、提案する中間コンテンツが、ビジュアル型言語とテキスト型言語の中間に位置づけられる特徴を有することをアンケートにより評価した。また、従来研究⁽²⁾では、ビジュアル型言語とテキスト型言語の学習の際に、学習者の生体情報（脳波、心拍、表情）を計測し、両言語の学習の際の生体情報に何らかの差異があるか否かの分析を行った。脳波を目的

変数、心拍と10種類の表情を説明変数として、重回帰分析を行ったところ、説明変数の係数の正負の符号に差異が見られた。

本研究では、従来研究⁽²⁾に引き続き、精神的ストレスの評価に向いているとされる心拍変動に着目し、ビジュアル型言語とテキスト型言語の学習の際のストレスの違いを評価する。

2 従来研究

Tóth ら⁽³⁾は、ビジュアル型言語とテキスト型言語間のギャップの存在を指摘した。この研究ではビジュアル型言語（MIT App Inventor 2）から Java Bridge Code Generator を知識伝達のメディエーターとして使用して、テキスト型言語（Android Studio）に移行する方法を検証した。この論文では Java Bridge Code Generator が、ビジュアル型言語とテキスト型言語の間のギャップを埋めるのに役立ったと主張している。

Weintrop ら⁽⁴⁾はビジュアル型言語から入ったが学習者とテキスト型言語から入った学習者がテキスト型言語に移行した後の知識移転の変化を実験で評価し、両者に有意差が無い事を示している。移行ではなくテキスト型言語のスキルを修得後の比較を行った。

我々は、以前の研究⁽¹⁾で、ビジュアル型言語の学習とテキスト型言語の学習の間に、我々が提案する中間コンテンツを用いた学習を挟む方がその後のテキスト型言語の理解度が高まるということを実証実験を通して示した。さらに、提案する中間コンテンツが、ビジュアル型言語とテキスト型言語の中間に位置づけられる特徴を有することをアンケートにより評価した。

また、我々は、脳波計を用いて学習者の学習状況を把握する方法の研究を推進中である。我々は、課題遂行中の脳波情報を取得し、難しい課題に取り組むと β/α の値が高くなることを示した⁽⁵⁾⁽⁶⁾。この研究の過程で、ビジュアル型言語 (Scratch) の課題

に取り組んでいる際の脳波とテキスト型言語 (C 言語) に取り組んでいるときの脳波に差異があることを確認した。具体的には、ビジュアル型言語に関しては課題の難易度が高くなっても β/α の値が高くなることはなかった。これによりビジュアル型言語とテキスト型言語の学習過程には異なる思考が行われている可能性が示唆された⁽⁷⁾。

3 実験方法

3.1 実験参加者

今回の実験の参加者は、湘南工科大学の4年生の7名である。彼らは同じ大学の同じ学部でプログラミング関連のコースを数年間学んだ経験を有する。彼らのプログラミングスキルはほぼ同等である。ただし、分析の後でわかったことだが、キーボードの入力技術に関して学生間で差があった。

3.2 実験に使用する Web サービス

ビジュアル型言語のプログラミングを行う際は、Google Blockly⁽⁸⁾ (図 1 参照) を利用した。このサイトにはパズルや迷路などの課題が掲載されているが、今回はテキスト型言語と内容を合わせるために、音楽の課題に取り組むこととした。また、テキスト型言語のプログラミングを行う際は、JSFiddle⁽⁹⁾ (図 2 参照) を利用した。このサイトはテキスト型言語である JavaScript 言語を実行できる統合環境である。リソース設定として Beepplay ライブラリを追加することによって Beep 音によって音楽を作ることができる。

3.3 実験に使用する問題

まず、練習を 1 回行い、その後、本実験を 2 回行った。具体的には楽譜を見ながらその楽譜の通りの音ができるようにプログラミングを行う課題を行った。それぞれの練習と実験で対象とした歌は表 1 の通りである。なお、これらの楽譜を印刷し、実験時に実験参加者に提示した。なお、実験で取り扱った楽譜に

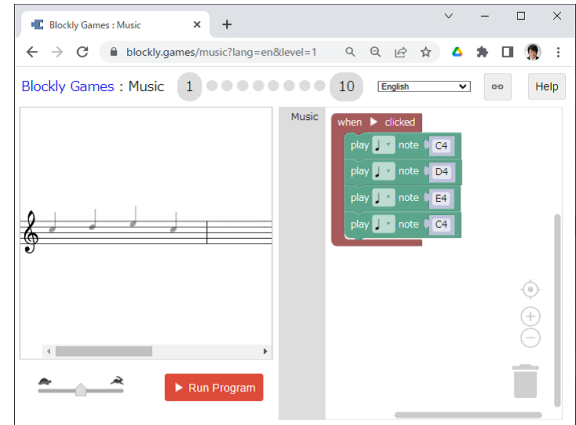


図 1: Google blockly (Music) の画面

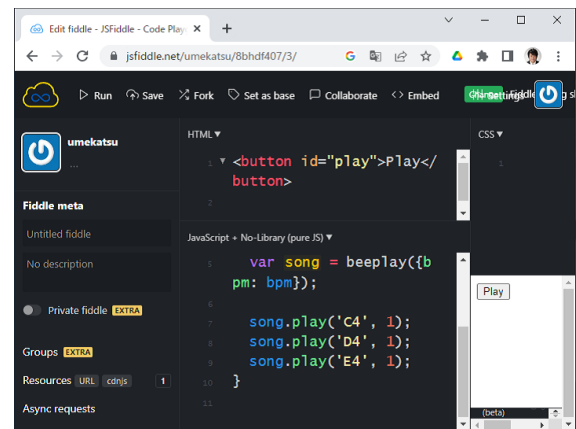


図 2: JSFiddle (Beeplay) の画面

は、音符の上に国際式表記 (C3, C4 など) が記入されている。図 3～図 5 にそれぞれの歌の楽譜の一部を示す。

表 1: 対象とした歌	
実験	曲名
練習	カエルの合唱
本実験 1 回目	メリーさんの羊
本実験 2 回目	ジングルベル

3.4 実験に使用する生体機器

3.4.1 心拍変動計

心拍変動の計測には Apple 社の “Apple Watch 8” を使用する。また、心拍変動の計測アプリとして W



図 3: かえるのうた (Froggy's Song) の楽譜



図 4: メリーさんのひつじ (Mary Had A Little Lamb) の楽譜

MS 社 (WMS, Inc) の“心拍変動の分析・心電図アナライザー”を使用する。このアプリは Apple Watch と連携して HRV (root Mean Square of Successive Differences: rMSSD) を計算することができる。Apple Watch 8 で心拍変動を計測する場合、腕に装着した状態で別側の手の指を Digital Crown (リューズ部分) に当てる必要がある。30 秒間指を当てたままにすることで心拍変動の指標の一つである rMSSD が計測される。

3.4.2 キーロガー

マウスを多用するビジュアル型言語と、キーボードを多用するテキスト型言語で、それぞれの入力デバイスの使用頻度等を観測するために実験用 PC にキーロガーを組み込み、実験中の入力デバイスの使用状況を計測した。

3.5 実験の流れ

図 6 に実験の全体フローを示す。図 6 に示すように、テキスト型言語から取り組むグループ (グループ A) と、ビジュアル型言語から取り組むグループ (グループ B) に分けた。心拍変動の計測は、3.4.1 節に示したように、指を Digital Crown (リューズ部分) に添える必要があるため実験の最中 (プログラミング中) に計測することができない。そのため



図 5: ジングルベル (Jingle Bells) の楽譜

に図 6 に示すように実験の前後 30 秒間で心拍変動を計測することとした。なお、キーロガーによる入力情報はそれぞれの実験中に取得している。

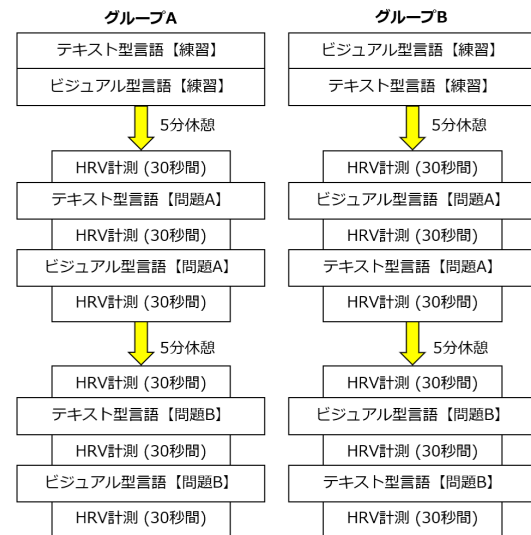


図 6: 実験の流れ

4 実験結果

グループ A (参加者 A~D) の心拍変動を図 7 に、グループ B (参加者 E~G) の心拍変動を図 8 に示す。実験の順番が心拍変動の変化に影響を与える懸念があったために、今回の実験ではグループ A とグループ B に分けて実験の順番を入れ替えた。図 7 と図 8 を見ると、実験参加者 C と G 以外は、どちらもグループの参加者もテキスト型言語の実験の前後で、心拍変動の値が上がっている。また、ビジュアル型言語に関しては、実験参加者 B 以外はどちらのグループの参加者も実験の前後で下がっている。一般的にストレスレベルが上がると心拍変動は下がり、

反対にストレスのバランスが保たれていれば心拍変動は上がると言われている。つまり、テキスト型言語では、多くの実験参加者はストレスバランスを保っているが、ビジュアル型言語では、多くの実験参加者はストレスレベルが上がっていると言えそうである。次節以降でこの結果が統計的に正しいと言えるか否かを検証する。

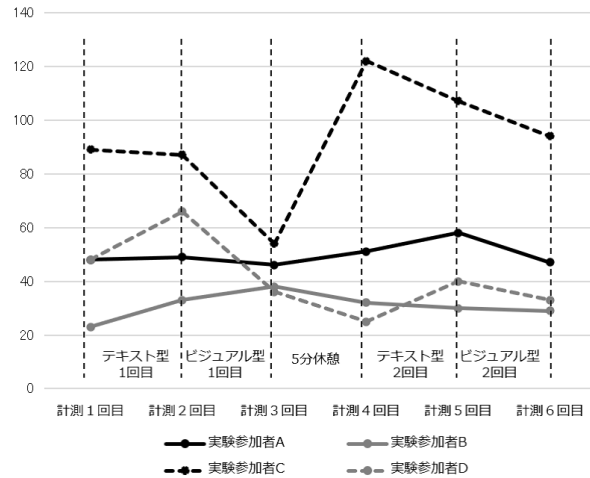


図 7: グループ A の心拍変動

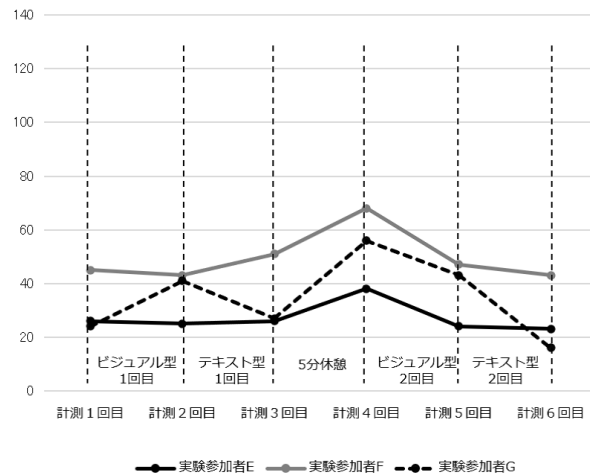


図 8: グループ B の心拍変動

5 分析

前節の図 7 と図 8 で示した通り、実験の順番が異なってもビジュアル型言語とテキスト型言語の実験の前後の心拍変動の変化の傾向に差異がなさそうである。よって本節以降の分析では、グループ A のテキスト型言語 (1 回目) の実験の前後の計測 (つまり計測 1 回目と計測 2 回目) と、グループ B のテキスト型言語 (1 回目) の実験の前後の計測 (つまり計測 2 回目と計測 3 回目) を統一的に扱う。他の計測に関しても、グループ A とグループ B の同種の実験の前後の計測を同じものとして扱う。

また、文献⁽¹⁰⁾によると、rMSSD の分布は正規分布には従わないが、rMSSD の値を自然対数変換をすることで正規分布に従うとされている。また文献⁽¹¹⁾によると、観測値の独立性、母集団分布の正規性、等分散性の前提があれば、サンプルサイズ (ただし $n \geq 2$) に関係なく t 検定などのパラメトリック検定が保証されるということである。さらに対応のある t 検定では等分散性の仮定は不要である。よって以降の分析では rMSSD を自然対数変換を行ったデータを使って対応のある t 検定を行っていく。

さらに文献⁽¹⁰⁾では、今回実験に用いたような 30 秒という短時間の計測による rMSSD などの HRV 指標の計測では、グループ間の統計量の検定を行う前に、同一グループ内で計測時間を延ばした実験結果と比較しなければならないとされているが、今回の分析では、30 秒以上の計測を行っていないため同一グループ内の検定は省略する。

5.1 母平均の差に関する t 検定

全参加者を対象に、テキスト型言語 (1 回目)、ビジュアル型言語 (1 回目)、テキスト型言語 (2 回目)、ビジュアル型言語 (2 回目) のそれぞれの実験の前後で心拍変動 (rMSSD) の自然対数変換値の平均値に差が出ているかを検定する。対応のある t 検定を行った。検定結果を表 2 と図 9 に示す。

表 2: 心拍変動の平均値の検定

実験	前/後	平均値	p 値
テキスト型 (1 回目)	前後	3.7229 3.7899	0.2604
ビジュアル型 (1 回目)	前後	3.7552 3.6761	
テキスト型 (2 回目)	前後	3.7443 3.6418	0.2811
ビジュアル型 (2 回目)	前後	3.9580 3.7209	

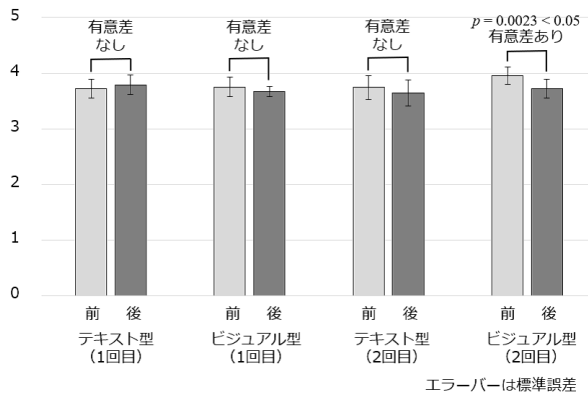


図 9: 心拍変動の平均値の検定

表 2 および図 9 に示した通り、ビジュアル型言語の 2 回目の実験の前後で心拍変動の値が有意に低くなった。その他に関しては統計的に有意な差とはなかった。この結果より、テキスト型言語の学習はストレスバランスには影響がないが、ビジュアル型言語の学習では、ストレスレベルが上がったと考えられる。

5.2 キーボード入力経験の違いを考慮した心拍変動の平均値の分析

前節では、実験参加者全員を対象に平均値の差の検定を行った。本節では、キーボード入力経験の違いに着目する。今回の実験では、各課題の実行中にキーロガーによりキーボードとマウスの入力を監視した。そのログを見ると、テキスト型言語の課題の際に、実験参加者 A, E, F, G の 4 名は Ctrl-C (コ

ピー) と Ctrl-V (貼り付け) を多用していることがわかった。逆に実験参加者 B, C, D の 3 名は、それらのシュートカットキーを全く使用していないことがわかった。このキーボード入力の経験が、今回のストレスレベルの変化に影響を与えているのではないかと考えた。Ctrl-C (コピー) と Ctrl-V (貼り付け) の多用者の検定結果を表 3 と図 10 に、Ctrl-C (コピー) と Ctrl-V (貼り付け) の不使用者の検定結果を表 4 と図 11 に示す。

表 3: Ctrl-C 多用者の心拍変動の平均値の検定

実験	前/後	平均値	p 値
テキスト型 (1 回目)	前後	3.6412 3.5944	0.3695
ビジュアル型 (1 回目)	前後	3.5337 3.6306	
テキスト型 (2 回目)	前後	3.6803 3.4324	0.1983
ビジュアル型 (2 回目)	前後	3.9857 3.6599	

表 4: Ctrl-C 不使用者の心拍変動の平均値の検定

実験	前/後	平均値	p 値
テキスト型 (1 回目)	前後	3.8318 4.0507	0.1066
ビジュアル型 (1 回目)	前後	4.0507 3.7367	
テキスト型 (2 回目)	前後	3.8295 3.9210	0.3391
ビジュアル型 (2 回目)	前後	3.9210 3.8024	

本結果より、テキスト型言語の課題解決時に Ctrl-C (コピー) と Ctrl-V (貼り付け) を多用しているためキーボードによる入力経験が高いと考えられる参加者は、ビジュアル型言語の課題解決時に心拍変動が統計的に有意に下がっていることが分かった。つまりマウス入力しかできないビジュアル型言語でストレスレベルが高くなってしまったということがわ

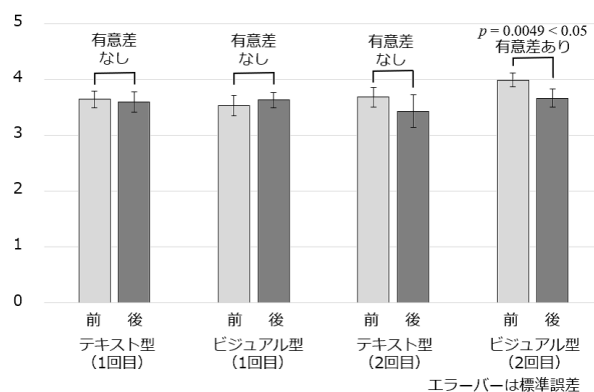


図 10: Ctrl-C 多用者の心拍変動の平均値の検定

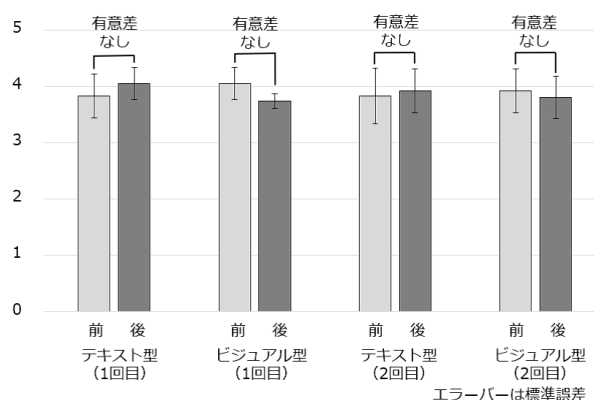


図 11: Ctrl-C 不使用者の心拍変動の平均値の検定

かった。

それに対して、テキスト型言語の課題解決時に Ctrl-C（コピー）と Ctrl-V（貼り付け）を全く使用していない参加者は、ビジュアル型言語の課題解決の前後で統計的に有意な差は見られなかった。

6 まとめと今後の課題

本研究では、ビジュアル型言語とテキスト型言語の学習の際に、両言語の学習の際の精神的ストレスに何らかの差異があるか否かを確認するために、学習者の心拍変動を計測した。ビジュアル型言語の学習時に心拍変動値が低くなる、つまりストレスレベルが高くなることを確認した。また、キーボード入力のログを取得することにより、キーボード入力に

慣れている被験者とそうでない被験者を別々に分析することによって、キーボード入力に慣れている被験者はビジュアル型言語の学習時によりストレスを感じていることを確認した。

今後、ビジュアル型言語とテキスト型言語間のギャップを埋める中間型言語を開発するにあたっては、今回の結論を踏まえて、様々な理解度や習熟度の学生に対してストレスを与えないような中間型言語を開発する必要があると考える。

研究倫理について

今回の実験は湘南工科大学研究倫理委員会の承認を得ている。また実験参加者から実験参加に関する署名を得ている。

謝辞

本研究は JSPS 科研費 JP22H01055, JP21K18535, JP20K03082 の助成を受けたものです。また、本研究の一部は、経営情報学会「ICT と教育」研究部会の助成を受けたものです。本研究成果の一部は早稲田大学理工総研プロジェクト研究「次世代 e-learning に関する研究」の一環として行われたものです。

参 考 文 献

- (1) 梅澤克之, 石田昂大, 中澤真, 平澤茂一. ビジュアル型言語からテキスト型言語への移行のための中間コンテンツの提案と評価. 電子情報通信学会 (IEICE) 技術報告, pp. 1-7, 2022.
- (2) 梅澤克之, 中澤真, 平澤茂一. ビジュアル型言語とテキスト型言語の学習時の生体情報の比較について. 電子情報通信学会 (IEICE) 技術報告, pp. 1-6, 2023.
- (3) Tomáš Tóth and Gabriela Lovászová. Mediation of knowledge transfer in the transition from visual to textual programming. *Informatics in Education*, Vol. 20, pp. 489-511, 2021.
- (4) David Weintrop and Uri Wilensky. Transitioning from introductory block-based and text-based

- environments to professional programming languages in high school computer science classrooms. *Computers & Education*, Vol. 142, , 2019.
- (5) Katsuyuki Umezawa, Tomohiko Saito, Takashi Ishida, Makoto Nakazawa, and Shigeichi Hirasawa. Learning state estimation method by browsing history and brain waves during programming language learning. *Trends and Advances in Information Systems and Technologies. WorldCIST'18 2018. Advances in Intelligent Systems and Computing, Springer, Cham.*, Vol. 746, , 2018.
 - (6) Katsuyuki Umezawa, Tomohiko Saito, Takashi Ishida, Makoto Nakazawa, and Shigeichi Hirasawa. Learning-state-estimation method using browsing history and electroencephalogram in e-learning of programming language and its evaluation. *Proceeding of the International Workshop on Higher Education Learning Methodologies and Technologies Online (HELMeTO 2020)*, pp. 22–25, 2020.
 - (7) Katsuyuki Umezawa, Yutaka Ishii, Makoto Nakazawa, Michiko Nakano, Manabu Kobayashi, and Shigeichi Hirasawa. Comparison experiment of learning state between visual programming language and text programming language. *2021 IEEE International Conference on Engineering, Technology & Education (TALE)*, pp. 01–05, 2021.
 - (8) Blockly games. <https://blockly.games/>. Accessed: 17 January 2023.
 - (9) Jsfiddle. <https://jsfiddle.net/>. Accessed: 17 January 2023.
 - (10) Leandro Pecchia, Rossana Castaldo, Luis Montesinos, and Paolo Melillo. Are ultra-short heart rate variability features good surrogates of short-term ones? state-of-the-art review and recommendations. *Healthcare Technology Letters*, Vol. 5, pp. 94–100, 2018.
 - (11) 水本篤. サンプルサイズが小さい場合の統計的検定の比較 —コーパス言語学・外国語教育学への適用—. 統計数理研究所共同研究レポート 238, pp. 1–14, 2010.

線形代数授業におけるオンライン問題の活用と学習分析

吉富 賢太郎

大阪公立大学 国際基幹教育機構

Use of Online Questions and Learning Analysis in Linear Algebra Courses

Kentaro Yoshitomi

Faculty of Liberal Arts, Sciences, and Global Education, Osaka Metropolitan University

大学初年次線形代数の授業において、動画開発による反転授業設計を 2014 年から行ってきたが、問い掛けの重要性から、LMS 上の CAS 利用のオンライン問題活用に基軸を移し、反転学習・授業内演習・復習用に活用している。オンライン演習を合格点まで繰り返し実施することにより、理解が深まることを目標としているが、学習分析を昨年度よりさらに多角的に進めることにより、個々の問題教材や授業設計の問題点を考察し、問題の活用方法や授業設計、学生の指導方法について改善点を考えたい。

キーワード： 線形代数, オンライン教材, 反転学習, 数学教育, 自動採点システム, 学習分析

1. はじめに

昨年度の第 2 回研究会において、筆者は、大学初年次の線形代数授業におけるこれまで行ってきた取り組みとオンライン問題の開発・運用と現状について報告した。以下にその概要を述べる。詳しくは、研究会報告⁽¹⁾を参照されたい。

工学部(旧工学域)大学初年次線形代数授業(80 人規模)向けのオンライン演習問題を、YouTube にて公開済みの解説動画の理解促進、反転学習の予習、対面授業における問掛けや授業内演習、自宅学習などを目的として通年の内容すべてについて開発し、LMS 上に設置、利用している。問題は、ほぼすべて、Moodle のプラグインである STACK⁽²⁾⁽³⁾によって開発されており、証明などについては、Moodle のドラッグ&ドロップ等を用いて構成している。なお、多肢選択問題も以前は自動生成と Moodle のランダム出題による疑似ランダム化を用いていたが⁽⁴⁾、現在は、STACK でフィードバック可能な実装方法を開発し利用している⁽⁵⁾。

演習問題の開発理念として、自分で考え、自分で解き進めるうちに、気付きや理解の深化がはかれるような概念理解が深まる問題を念頭においている。単なる計算問題はほぼ皆無で、計算過程の入力や、問題文を読

んで考える必要のある問題を中心としている。また、行列の基本変形などの計算問題も、Wolfram Alpha(あるいは、chatGPT)などに質問して答が得られる問題もあるが、入力する手間よりも自分で考えて答えた方がはるかに早く効率的であるような問題としており、これらは、チーティング耐性を持つ問題として開発方針の 1 つである。

このようなオンライン問題を学生に実施させる動機付けとして、達成度(合格点(10 点中 8 点)に達した問題の割合)で、平常点(40~50 点)のうち 10 点程度成績に加味したり、Moodle の機能であるデジタルバッジの付与を利用した。しかし、対面になってから特に、オンライン演習の実施率はあがっていない。

オンライン演習問題の達成率と紙の小テストや期末テストの間に弱い相関は認められるものの、目標が達成されているとは言えない状況である。

以上が去年(まで)の報告の概要である。

今回は、上記で活用した授業において、対面授業なのに欠席する学生数がコロナ禍前よりむしろ増えていることやオンライン演習を実施しない学生が相当数いる点について、学習分析により分析・検討した内容について報告する。ここで言う学習分析は、2021 年度と 2022

学籍番号	氏名	E1-1-1	E1-1-2	E1-1-3	E1-2-1
AAA0000	山田花子	10.0	9.0	1.0	6.0
AAA0001	大阪太郎	0.0	0.0	0.0	0.0
AAA0002	京都次郎	10.0	9.0	2.0	8.0
AAA0004	青森三郎	10.0	8.0	1.0	10.0

図1 オンライン演習データ見本図

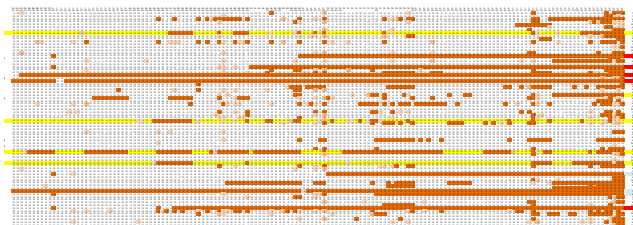


図2 演習実施の視覚化と成績 (2021 年前期)

年度の通年の内容について、オンライン演習やオンライン小テストの学習データ (Moodle の評定者レポート) を分析し、また、個別の問題について、学生の全般的な実施状況やさらに、個々の学生の実施状況の分析を指す。合わせて、学生への面談やアンケートを通じて、オンライン問題の構成や授業モデルについて、何が問題であったか分析し、オンライン問題を活用した効果的な授業設計について検討する。さらに、現在進行中の 2023 年度授業において、検討内容を反映できている範囲での実施状況について報告する。

2. 学習分析

2.1 実施状況の視覚化と大域分析

Excel における条件付書式を用いて、オンライン演習の最終得点 (すべてデフォルトの 10 点満点で統一している) について、色分けを行い、オンライン演習問題の実施状況の視覚化を行った。テーブルの構造を示すサンプル図を図 1 に、2021 年度と 2022 年度の前期の実際のデータを図 2,3 に示す。図 1 は、図 2,3 の右上を拡大するとおおよそこのようになっているという意味である。少し重いが、投稿したままの PDF であれば、拡大して実際のデータを見ることができる (個人情報 は削除してある)。

この視覚化により、次の 2 つの方向について、分析することができる。

- 横方向 (学生毎) の分析
- 縦方向 (問題毎) の分析

横方向、つまり、学生毎の分析では、まったく手をつけていない学生や最初はやっていたが、途中でやらな

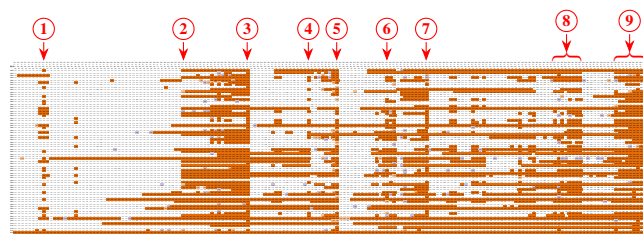


図3 演習実施状況の視覚化と成績 (2022 年前期)

くなった (ついてこれなくなった) 学生が受かびあがる。この分析は、授業期間が終了してからでは、実は遅い。授業開始から、分析内容を随時監視すべきであり、あやしい学生には早い内から声かけすることが重要である。例えば、やらなくてもできる学生であれば、声かけに対してきちんと回答してくると思われるので、できるだけ早いうちから声かけすることが望ましいと考える。

一方、縦方向、つまり、問題毎の分析では、問題の特徴を分析することが可能であろう。ほぼ誰もができていない問題、あるいは、できている学生が半分くらいしかない問題など、問題があるかもしれない問題の発見が可能である。また、現在はオンライン演習問題は、階層や分類を与えずに均等な展開を行っているため、この分析によって、逆に分類ができる可能性がある。

リアルタイムな分析により、バグがある問題の早期発見に繋がる。学生がバグに気付いて連絡してくれる場合があるが、できれば、教員側の考察で発見して、修正するのが望ましいのは言うまでもない。

また、認知的負荷が高い、あるいは、学生が苦手な形式や内容でいきなり取り組むのは難しい問題である可能性もある。「数学者が当たり前と考えることは学生には当たり前ではない」という法則はさまざまところで顔を出す。

このような問題を縦方向の分析で発見した場合、個別の問題検証、つまり、より局所的な分析が必要となる。

2.2 学習分析: 局所的な分析

ここで言う局所的な分析とは、前節のような視覚的に捉えられた情報に基づいて、個々の学習者や個々の問題に対して行う分析を指す。

まず、問題のある学習者を発見した場合、通常、LMS では、その学習者に限定した学習に関するデータを表示することができる。Moodle では、学習者のアクセス分析から、「詳細レポート」、「アウトラインレポート」

「評価概要」、評価の「シングルビュー」などを用いて学習状況を把握することが可能である。

また、特定の問題の特定の学生の解答を分析することが、最も局所的な分析であるが、LMS から得られるデータとしては、取り組み時間、取り組み回数、得点の推移、解答内容があげられる。

本稿では詳しくは述べないが、解答内容の分析は、誤答分析による誤答パターン解析に有効で、誤答パターン解析は、解答に応じたフィードバックに欠くことができない重要なステップである。通常は、フィードバックを完備していてもなお、想定外の解答が出た場合にさらに誤答を分析し、フィードバックを充実化することにより、学習教材としての問題教材の品質が向上する。

また、学生への面談も非常に有効である。昨年度より、大福帳.js⁽⁶⁾ を利用して毎回のコメントを学生に記入してもらっているが、その中でヘルプ信号を発する場合もあり、今年度、すでにそれに対応して、数学相談室での個別ケアをしている。

一方、問題については、視覚化された情報から、誰も達成されていない問題がいくつか受かびあがってくる。図 3 において、①～⑧で示した部分である。本稿では、これらの問題について、以下の節で簡単に説明する。

2.3 問題の分析例

①については、問題にバグがあり、多肢選択問題の選択肢が、プログラムの typo により、確率的に選択肢不足のエラーが発生していた。このため、全員ではないが、この段階で目立って未達成の問題となっていたのである。自習教材としてのバグは非常に致命的であり、教員が想定する以上に留意すべきポイントであることが、これまでの経験からわかっているが、問題開発に傾注している段階ではなかなかバグの発見に神経が回らず、このようなケースが当座発生することになる。回避する方法としては、STACK の問題編集において行える「問題のテストとデプロイ」を使用することである程度回避できる。今年度は、ほぼすべての問題において必ずこのデプロイによってパターンをあらかじめ生成しておいて確認するようにしている。

②は、図 4 の問題 (E2-4-01) である。この問題は、説明をじっくり読んで、行列の写像の定義を理解するための問題である。非常に基本的で、しかしながら、後期に至るまでこの概念で混乱を生じる学生が多い。例え

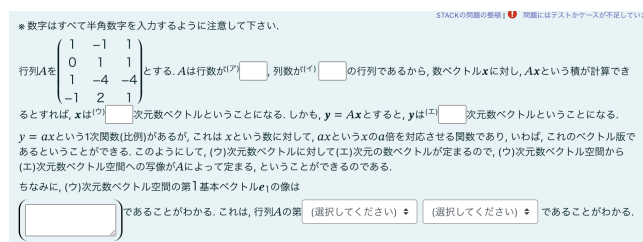


図 4 E2-4-01

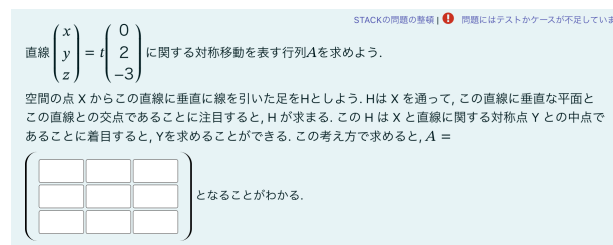


図 5 E2-5-05

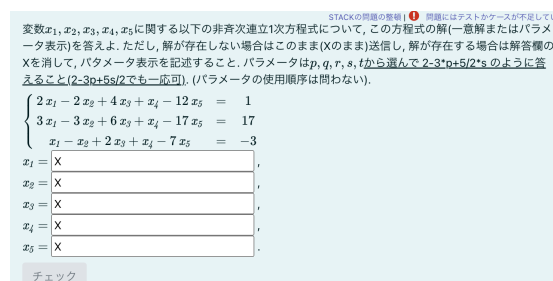


図 6 E3-4-14

ば、 3×4 行列の定める写像が $R^4 \rightarrow R^3$ なのか $R^3 \rightarrow R^4$ なのかがわからないのである。授業における説明でも、 3×4 行列を左からかけることができるベクトルは何次元の数ベクトルか、という問掛けをし、本演習問題でも同じ問掛けをしているのであるが、行列を写像と考えるところで認知的負荷がかかっていると考えられ、首をかしげる学生が多い。数学教育としてよりふみこんだ研究が必要なテーマの 1 つとなっている。

③は、図 5 の問題 (E2-5-05) である。この問題の場合、考え方がわからず手をつけられなかった学生が多かったと思われる。しかし、問題の難易度としては、高等学校の昔のカリキュラムの範囲の問題であるから、それほど難しくない。考えさせるという目的では向いている問題であるが、概念理解を深める問題として適切であるかどうかは検討の余地があると考えられる。

⑥は、図 6 の問題 E3-4-14 である。ただし、この問題では、このタイプの問題が 5 題構成されており、時間的に全部やりきることができなかった学生が多かったと考えられる。課題として、必ず実施させる場合には、各問題の構成問題数を適度に調整することが肝要だ。

⑧、⑨は、最後の節で行列式に関する問題の後半である。

本学では、行列式は抽象的な定義を教科書としては用いており、学生の認知的負荷は明らかに高い。特に、多重線形性の概念で戸惑う学生が多い。そのため、2次行列式、3次行列式で、できるだけ多くの実例に問題で触れてもらうことで概念理解に至るように問題開発しているが、自習だけで進めるにはかなり困難であると考えられる。

今年度は、それを見越して、2次行列式の節 (§4-1) を先行してやるように学生に声をかけているところであるが、今のところ実施している様子はない。

このような、概念的理解の困難性と、それまでの内容が押している関係で、最後は息切れ状態になっているのが⑧、⑨であると考えている。

その他については、本稿では略すが、多くの、あるいはほぼすべての学生がつまづいている場合、その問題が「公文式」的にこなす問題としては、不適切であり、授業形態や授業設計、利用目的（評価が課題か学習か）に応じて使い分けの必要があることが明確になってきた。

今後、このようなつまづきの問題のうち、不適切な問題は避け、あるいは、ヒントや説明の方法、出題形式に改良を加えることによって設置方法の改良が必要である。

3. Moodle への配置と授業設計の検討

本節では、前節で行った分析に基づき、オンライン問題の Moodle への配置と授業設計について検討する。今年度は、この検討に基づいた試みを行っている。

3.1 Moodle への配置

反転授業を目的としたオンライン演習問題を開発した当初は予習用問題、その他の演習用問題として分類していたが、予習用問題を実施完了できる学生が極端に少なく、最近では、均等に問題を配置するようにしていた。これは、学生に、自分でできる問題をできるだけやってもらう、というねらいであり、特にコロナ禍ではそのようにするのが妥当と思われた。

その後、対面に移行していく過程（2021 年度～2022 年度）で、次第にオンライン演習問題の実施状況が悪くなった。このような状況で前節までのような学習ログの分析を行った結果や、大福帳による学生の声などから

判断して、予習用の問題は一定程度設け、授業内演習用、復習用に分類することがやはり妥当ではないかと考えるに至った。これは、昨年度の研究会でも示唆されたことである。

また、授業内演習や復習用の問題は、繰り返し実施することが望ましい。これは、形成的評価を目的としたものであり、数学教育で用いられる APOS 理論の Actions と Process を ICT で支援するものと考えている。繰り返す内に認知的負荷が下がり、概念的に自明と思える一種の内化が、学生に起こることを期待する。

ただし、学生によっては、解ける、あるいは理解できるという問題を解かせる必要はない。そこで、小テストを設置し、できなかった問題だけを繰り返し実施してもらうのがよいと考える。これは、自習教材のデザインとしてのプレテスト-学習-ポストテストという設計を、小テストに2回以上の受験可能回数を設定することにより、エミュレートする問題配置と言える。

3.2 授業設計の検討

授業設計について、授業の基本的なスタイルに応じて検討したい。

オンライン授業 コロナ後では、完全なオンライン授業になることは今後はないと考えられるが、オンデマンドオンライン授業用の教材として、臨時的に、あるいは、学生の自習教材として活用は可能である。その場合、自習という性質から、説明の必要なものには何らかのマークをつけ、基礎的で自習で進めやすい問題に取り組ませる必要があると考えられる。

ガイド付きの問題で分析から多くの学生が簡単にできているような場合は予習、授業でも説明を補足した方がよいような問題は、授業内演習として利用するのが妥当であろう。

対面授業 通常の対面授業においても、大福帳における学生のコメントを見ると、予習をすることで、授業で聞きたいポイントを絞ることができたり、授業で説明を改めて聞くことにより理解が深まったという意見が多く見られる。このような点から、多くの学生が躓くことなく解答可能ないくつかの問題を予習教材として設置することができる。今後は、予習したからわかったという学生が具体的に実施した問題を特定し、予習用問題として、配置もしくは授業で指示するのがよいかもしれない。ただし、反転授業ではないので、予習してこ

なかったことで不利にならないようにするという配慮も必要となり、通常の講義の中で簡潔かつ平易に説明することが効果的であろうと考える。

また、授業内の演習教材については、説明の必要な問題を授業内で解説し、その場で解答させ、できなかった範囲は復習に回す、という利用方法が考えられる。大阪府立大学において運用していた MATH ON WEB を活用した授業モデル⁽⁷⁾を STACK を使って実現したモデルとすることができる。

さらに、この授業モデルにおいては、次回授業の冒頭で紙の小テストを行っていたが、オンライン問題に切り替えられる問題も多くあり、それらを使うことによって、授業での解説や演習時間をより多く確保できる。ただし、この場合も学生の自宅学習時間が過剰にならないようにすることも留意したい。

反転授業 反転授業は、予習を学生にしてきてもらうことを前提にする授業モデルであるが、数学の場合、「予習しても理解できない」という現象が発生する。知識ベースの授業や日常に密着した科目においては、「理解できない」という状況はまず起こらないが、特に代数系の科目では、概念が理解できないという現象が発生する。例えば、後期の「抽象ベクトル空間の定義」を予習させることは困難である。この点は、微積分学を含めた他の科目と異なる点であると言える。後期授業開始前に Moodle のアナウンスで予習をよびかけたことが何度もあるが、うまく行ったことはない。前期までの範囲であれば必ずしも、無理ではないが、空間図形の平面の方程式や行列が高校カリキュラムからはずれている現状では、上述の「予習をできる範囲でしてもらう対面授業」が線形代数に関しては、適しているのではないかと考えている。

ただし、問題や動画の改良によって、今後は少しずつ反転形式に近づける研究を進めていきたい。

3.3 現在の実施状況

最後に、2023 年度前期における授業設計と運用状況・学生の実施状況について原稿執筆時点での状況を付記しておく。発表では、最新の状況について報告する予定である。

今年度は、3.1 で述べたように、オンライン演習の実施は任意とし、細かくステップに分けて、オンラインテストを設置している。オンライン演習をやらなくても、

+ S2-4.1次写像と行列

- + 概要・解説
- + ★ --- 行列の定める写像の定義 --- ★
- + E2-4-1a 行列の定める写像の定義
- + E2-4-1b 行列の定める写像の定義と像
- + E2-4-1c 行列の定める写像と基本ベクトルの1次結合の像(ベクトルの対応から行列を選択)
- + S2-4-1 5/27 5:00 まで 20分 x2 ≒ 7
- + ★ --- 線形性と1次結合 --- ★
- + E2-4-2a 線形性と1次結合の像(基本ベクトルの像から計算)
- + E2-4-2b 1次結合の係数を求めて像の計算
- + E2-4-2c 1次変換による平面全体の像
- + E2-4-2d 平面の1次変換による直線の像の考察
- + E2-4-2e 空間全体の1次変換による像を考える
- + E2-4-2f 空間の一次変換による平面の像
- + S2-4-2 利用制限 次の条件に合致しない限り利用できません。活動「S2-4-1」が完了マークされた場合
5/27 5:00まで 20m x2 ≒ 6
- + ★ --- 線形写像を表す行列 --- ★
- + E2-4-3a 数ベクトル空間の間の多項式で書ける写像が線形であるための条件
- + E2-4-3b 基本ベクトルの像を表す行列
- + E2-4-3c R2の2つのベクトルの像から表す行列
- + S2-4-3 利用制限 次の条件に合致しない限り利用できません。活動「S2-4-2」が完了マークされた場合
5/27 5:00 まで 15点満点 30m x 2 ≒ 9
- + まとめ2-4

図 7 2023 年度配置例

オンラインテストを完了すれば一定の平常点になるという設定である。

オンライン小テストは、すべて 2 回以上受験可能であり、1 回受験して、できなかった問題をオンライン演習でじっくり取り組んで再度受験ということが可能である。

また、各節毎に 2~4 程度のオンライン小テストがあるが、順に完了しなければ先に進めないように、利用制限を Moodle の活動完了の機能を用いて設置している。最後に行くほど配点が高くようにしておけば、学生はすべてを実施することが重要になる(図 7)。

図 8 に小テストの現在までの実施状況を示した。赤い部分は 2 点以下で、実質未受験と考えられる部分である。中央の赤い縦の部分は、必須ではないので後回した部分である。学生によっては、先まで進んでいる学生もいるのがわかる。

また、特定の学生 2 名は一切活動がないが、それ以外で見ると、1 名かなりの重症と言える状態の学生がおり、他に 5 名、個別に指導が必要と思われる学生がいるこ

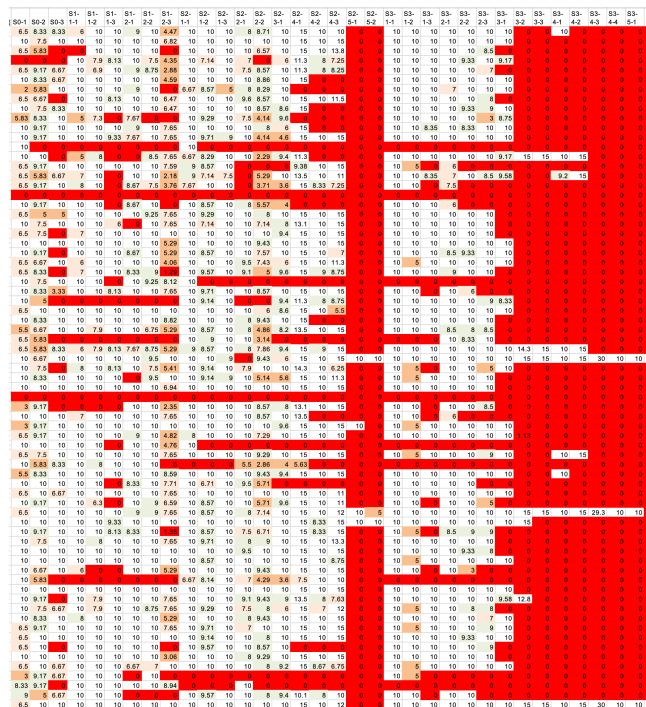


図8 2023年度オンライン小テストの状況

とがわかる。

このように、小テストでの状況にしぼって分析することで、早期に対処することが可能である。実際、初期にメールで声かけした何人かのうち、一名はその後、授業後や数学相談室に来て質問をしに来るということがしばしば見られ、個別対応の効果はあると考えている。

上述の学生群には早い段階で連絡はしたが反応がなかった。今後、再度連絡して対応していく予定である。

4. まとめと課題

昨年度に続き、大学初年次数学(線形代数)において、オンライン演習を活用し、学生の概念理解の促進を目指している。これまでの学習ログデータの分析結果から、学生毎の要指導学生の検出、要改訂の問題の検出、問題の目的別分類への利用への可能性について検討した。

今年度は、上記の分析を踏まえて、自習教材のプレテスト-学習-ポストテストの考え方を適用し、

- ・小テストができれば演習は不要(負担軽減)
- ・小テストに合格すれば次に進める(ゲーム性)
- ・小テストのために演習を反復(形成的評価)

のサイクルを目指して運用している。現在のところ、格段に指導が必要な学生を数名検出しており、今後の指導に役立てたいと考えている。

今後は、学生からの質問の多い問題は、時間のかかる

問題やじっくり取り組むべき問題として課題用とするなど、より適切な分類と問題配置を行い、反転学習用の予習問題としての分類も行って、学生・教員の両者にとって、負担は少ないがより効果的な教材として整備していくことが課題である。

謝辞

本研究は JSPS 科研費 21H00921 の助成を受けたものです。

参考文献

- (1) 吉富 賢太郎, “線形代数における形成的評価を目的とするオンライン演習問題について”, 教育システム情報学会 2022 年度第 2 回研究会報告, pp.49–56, (2022).
- (2) STACK ホームページ <http://stack.bham.ac.uk/>
- (3) https://github.com/maths/moodle-qttype_stack (2023 年 6 月 15 日確認)
- (4) K.Yoshitomi, “Generation of Abundant Multi-choice or STACK Type Questions Using CAS for Random Assignments”, in “Mathematical Software – ICMS 2018”, Springer, pp.492–497, (2018).
- (5) K.Yoshitomi, “Multiple-choice questions using STACK with partial score and feedback”, ATCM 2022 Proceedings, pp.333–342, (2022),
- (6) 大福 帳.js <https://goose.cite.tohoku.ac.jp/daifukujs/> (2023.6.15 確認)
- (7) K.Yoshitomi, M.Kawazoe, “E-learning/e-assessment systems based on webMathematica for university mathematics education”, MSOR Connections Vol15(2), pp.17–24, (2017).

ドローイングストロークの線形回帰に基づく

パース技術評価に関する基礎的検討

Basic study on evaluating perspective techniques based on linear regression of drawing strokes.

Kanato SUGII^{*1}, Takashi NAGAI^{*2}, Mizue KAYAMA^{*1}

^{*1} Graduate School of Science and Technology, Shinshu University

^{*2} Institute of Technologists

概要：本研究の目的は、ドローイングスキルの習熟度を定量的に評価するための指標の具体化である。観察ドローイングにおいて、奥行きを捉える技術（パース）は重要とされている。本研究では、独自開発した遠隔ドローイング学習支援システムに蓄積されたドローイング作品に含まれる、ストロークの座標データからパースの良不良を評価するための基礎的検討を試みた。具体的には、ドローイング作品のストローク群からパースに関係するものを抽出し、線形回帰により消失点を推定した。本稿ではまず、観察ドローイングにおけるパースと、本研究で扱うドローイングデータについて述べる。次に、線形回帰に基づくパースと消失点について示し、学習者と指導者の結果を比較する。その上で、推定消失点のドローイング習熟度評価指標としての可能性と、自動評価の方法について考察する。

キーワード：ドローイング，ストローク，パース，消失点，線形回帰，習熟度評価，スキル獲得支援

1. はじめに

ドローイングは美術教育における基本的な技法として位置づけられ、入門者が最初に学ぶべき内容とされている(1)(2)。観察ドローイングにおいて、対象物（以下、モチーフ）の奥行き表現する技術（以下、パース）は重要なものとされている。しかし、パースの技術を初学者が身につけるためには多くの時間を要する。また、初学者が描画中に自身のパースの良不良を認識することも難しい。そこで、本研究ではパースの習熟度を示す定量的な指標を探究することで、初学者におけるドローイングスキルの獲得支援を目指す。永井らの研究(3)によって、ドローイング作品を構成す

る線（以下、ストローク）のキャンバスに対する座標情報の記録が可能となっている。本研究では、描かれたストロークの座標情報に基づく解析から、ドローイング作品に用いられているパースを可視化し、ドローイングスキルの定量的な評価が可能となる仕組みの実現を目指す。その基礎的検討として、本稿では、ドローイング作品におけるパース推定手法を検討し、学習者と指導者の比較から、ドローイングの習熟度を定量的に比較する長指標としての可能性を考察する。

2. ドローイン作品のストロークデータ

2.1 遠隔ドローイングシステムの概要

ドローイングの学習を個人で行う場合、自分の癖や弱点に自ら気づくことは難しい。そのため、美術スクールに通い指導を受けることが多い。しかし、一対多の授業形態では指導者が学習者一人ひとりのドローイングをすべて把握することは難しい。そこで永井らは、ネットワーク環境を用いた美術入門者のためのドローイング学習支援システムを構築した(3)。このシステムを利用することで、学習者は時間と場所に制約を受けることなく指導者からアドバイスや評価を得られるようになることが期待される。このシステムは、学習者がドローイング作品を描く過程で用いたすべてのストロークの幾何的情報を記録している。

2.2 ドローイングのプロセスデータ

永井らのドローイング学習支援システムでは、筆記具としてデジタルペンを、学習プラットフォームとして Plone CMS（以下、CMS）を利用する。デジタルペンが記録したドローイング過程のストロークデータをドローイングプロセスデータと称する。このシステムではドローイングプロセスデータを CMS に蓄積する。ここで、ドローイングプロセスを再生するためのツールであるドローイングプロセスビューアの例を図 1 に示す。これはドローイングプロセスを時系列解析したグラフエリアと再生エリアから構成される。学習者および指導者は、ここで再生されるドローイングプロセスを確認しながら学習や指導を行う。

学習者や指導者はドローイングプロセスビューアを用いてドローイングプロセスを参照・閲覧できる。これにより、オンラインであってもオフラインの学習に近い体験が得られる。時間と場所に制約されずにドローイングプロセスを再生可能なため、指導者は全学習者のドローイングプロセスに対する指導が可能となる。また学習者も他者のドローイングプロセスを参照・再生可能となるため、自身のドローイングプロセ



図 1 ドローイングプロセスビューア

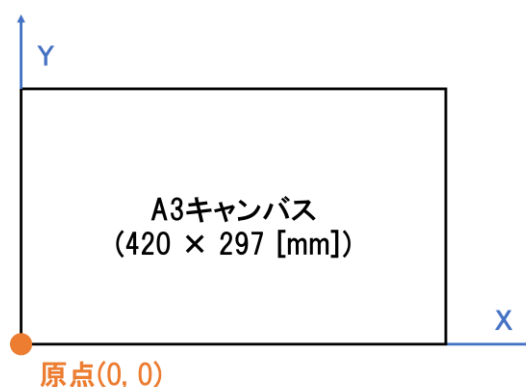


図 2 A3 サイズのキャンバス

スとの違いや優れた描画技法を観察することができる。

2.3 ドローイング作品のストロークデータ

ドローイング学習支援システムに蓄積されているドローイングプロセスデータには、ストロークの描かれたタイミングや筆圧、キャンバスに対する座標の情報などが含まれている。これらの情報の中から本解析では、キャンバスに対する座標情報を用いる。座標情報は、図 2 に示すキャンバス上のある 1 頂点を原点とした x 座標 $X(0 \leq X \leq 420 [\text{mm}])$ と y 座標 $Y(0 \leq Y \leq 297 [\text{mm}])$ で表される。

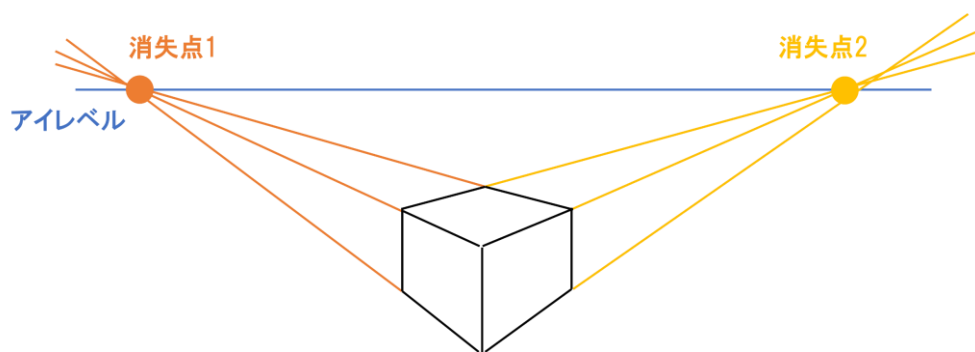


図 3 2 点透視法による消失点

3. パースの可視化

3.1 パース

パースとは、英単語の **perspective** の略称であり、ドローイングにおいては遠近法を用いた描画手法を指す。パースは、平面上でモチーフの奥行きを感じさせる表現技法である。パースには様々な種類があるが、本稿では線遠近法によるパースを対象として解析することとする。線遠近法は、「人の目には奥へ伸びる平行線が一点へ収束して映ることに起因する人の奥行知覚を利用し、収束する平行なストロークを描くことで遠近感をおこさせる技術である」(4)と定義されている。ストロークが収束する一点を“消失点”と称する。線遠近法には、モチーフに対する消失点の数に応じて3種類(1点透視法、2点透視法、3点透視法)存在する。2点透視法を用いて紙箱のような単純な直方体をモチーフにドローイングを描く場合には、図3のように1つの消失点に対して3つの辺が収束する。全ての消失点はアイレベルと呼ばれる水平線上に位置することが理想とされている。

3.2 解析対象

3.2.1 対象データ

永井らの遠隔ドローイング支援システムには841件のドローイングプロセスデータが蓄積されている。これらのデータは2012年～2022年までの11年間、美術専門学校でのデジタルドローイングクラスにおいて生成された。デジタルドローイングクラスでは、規定時間以内に指定されたモチーフに対してデジタルペンでドローイングを行う。規定時間は20分とされ、指定モチーフはクラス前半が紙箱、後半が紙袋とされた。

本稿では予備解析として841件のデータの内、学習者3名(以下、A, B, C)と、指導者1名の計4名のドローイングプロセスデータを解析対象とする。上記の3名はドローイング指導者により、デジタルドローイングクラスでの成長の度合いが高いと判断された学習者である。なお、AとBは完全な初学者であり、Cはドローイング学習の経験を有する学習者である。

ドローイングプロセスデータのモチーフは、紙箱

と紙袋である。予備解析ではパースの可視化が容易である紙箱をモチーフとした12のドローイング作品(以下、紙箱ドローイング)を対象に解析する。解析対象とする12種のデータは、解析対象者4名それぞれに、異なる紙箱をモチーフとした3回分のプロセスデータである。なお、学習者については、1回目の紙箱ドローイングが初回学習であり、2回目、3回目の紙箱ドローイングはそれぞれ前回のドローイング学習から2週間後に描かれたものである。解析対象とした12種の紙箱ドローイングのアウトライン描画段階の終了時刻は、浅井らの先行研究(5)によって定義済みである。

3.2.2 アウトライン描画段階

本研究では、ドローイングプロセスデータのうち、描画開始からアウトライン描画段階(5)までの区間を対象にパース解析を行う。アウトライン描画段階とは、「ドローイングにおいて、最初のストロークを描き始めてから1番最初のトーンが描かれるまでの段階」のことを指す(5)。トーンとは、モチーフの明暗や質感を描画するために引かれる複数のストロークを指す。

以下に、アウトライン描画段階までのプロセスデータを用いる理由を示す。

- ① アウトライン描画段階は、モチーフの概形を捉える段階であり、パースの正確さはこの段階でほぼ決定する。
- ② トーンが描かれる前のドローイング作品から、パースを構成しているストロークを判別することは可能である。

図4では、アウトライン描画段階終了時のドローイングと指導者の1個目の紙箱ドローイングの完成

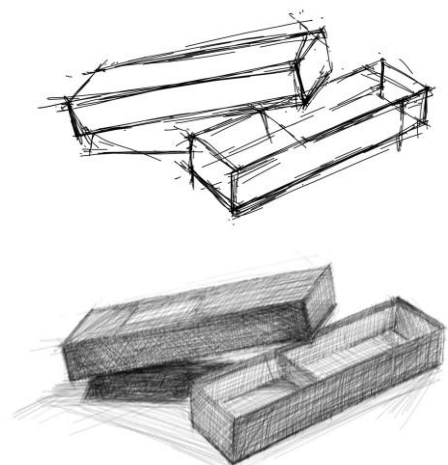


図4 アウトライン描画段階終了時(上)と完成時(下)の紙箱ドローイング

作品を比較する．アウトライン描画段階終了時と完成時において，モチーフのアウトラインに大きな差はないことが分かる．また，アウトライン描画段階終了時にはトーンが描かれていないため，パースを構成しているストロークを見分けやすいことが分かる．

3.3 解析手法

解析は以下の手順でおこなう．

- I. ストロークをモチーフの辺ごとに分類する．
- II. 分類されたストローク群の座標データを用いて線形回帰をおこない，求められた傾きと切片から辺の延長線を可視化する．
- III. 可視化された辺の延長線を，同じ方向に向いている 3 本 1 組で分類し，消失点方向における交点を求める．この 3 本 1 組の辺を“消失点構成辺”と定義とする．
- IV. IIIで求めた交点の重心を消失点として推定する．なお，交点が 1 つであった場合には，その 1 点を消失点とした．

線形回帰には，Python の `scikit-learn` を用いた．4 名の対象者が描いた紙箱ドローイングは，B の 3 回目は 1 点透視法で，それ以外は 2 点透視法で描かれていた．このことを踏まえて，パースを以下の 3 つの項目で定量評価する．

- ① 消失点構成辺の交点の数：3 つの交点が存在している状態を理想とし，交点の数が多いほど理想に近いとする．
- ② 消失点構成辺の交点同士の距離の総和：交点同士の距離の総和が小さいほど理想に近いとする．
- ③ 2 つの消失点の y 座標の距離の差：2 点透視法の場合に，2 つの消失点の y 座標の距離の差が少ないほど理想に近いとする．

ここで，②については，①の交点が 2 つ以上存在することを前提とする．また③は，①の交点が 1 つ以上であることと，解析対象のパースにおけるアイレベルがキャンバスの底面に対して水平であることを前提とする．このことから，図 4 に示したような 2 つの重なっている紙箱をモチーフとしたドローイングでは，紙箱の底面が地面に接しているデータを評価対象とした．

3.4 解析結果

解析結果の例として，学習者 A の 1 回目のドローイングと，指導者の 2 回目のドローイングのパースを可視化した成果を図 5 と図 6 に示す．図 5 は，それぞれのドローイングのアウトライン描画段階終了時の消失点構成辺を示している．それぞれの消失点構成辺を青色とオレンジ色で色分けした．どちらのドローイングも 2 点透視法で描かれている．図 6 は，箱の各辺を構成しているストローク群の座標情報に対して線形回帰をおこなった結果である．図 6 中の 4 つの黒丸はそれぞれ，A3 キャンバスの 4 つの頂点を表している．図 6 上に示す学習者 A の結果から，オレンジ色の消失点構成辺は収束傾向にあるが，青色の消失点構成辺は収束傾向にないことが分かる．一方で指導者は，オレンジ色と青色の消失点構成辺共に，一点に収束している．

解析対象である 12 個の紙箱ドローイングにおいてパースを定量評価した結果を表 1 に示す．定量評価の項目は，3.3 節で述べた 3 項目である．例えば，学習者 A さんの 1 回目の紙箱ドローイングについては，2 点透視法で描かれているので 2 組の消失点構成辺に対する解析結果が示されている．消失点 1 の方向の消失点構成辺は，交点が 2 個存在し，その 2 点間の距離は 1999[mm]であった．消失点 2 の方向の消失点構成辺は交点が存在しなかったので 0 個とした．また，交点が 0 個であるため，交点の距離の総和による評価と，消失点 1 と消失点 2 が同一のアイレベルに位置しているかの評価結果は記号「-」で示した．

3.5 考察

3.3 に示した各定量評価項目の有用性を考察する．まず，消失点構成辺の交点の数は，消失点構成辺がどの程度 1 つの消失点に集まっているかを示している．つまり，交点の数が多いほど消失点を意識したドローイングができていると解釈できる．よって，交点の数

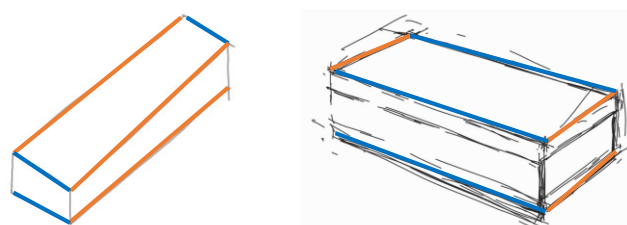


図 5 学習者 A の 1 回目の紙箱ドローイング（左）と指導者の 2 回目の紙箱ドローイング（右）

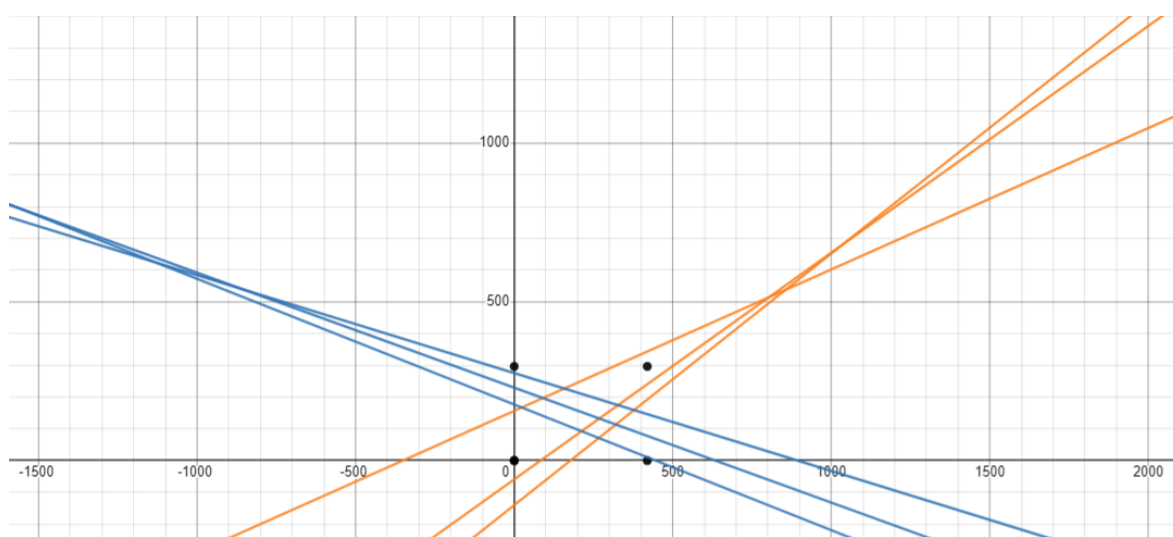
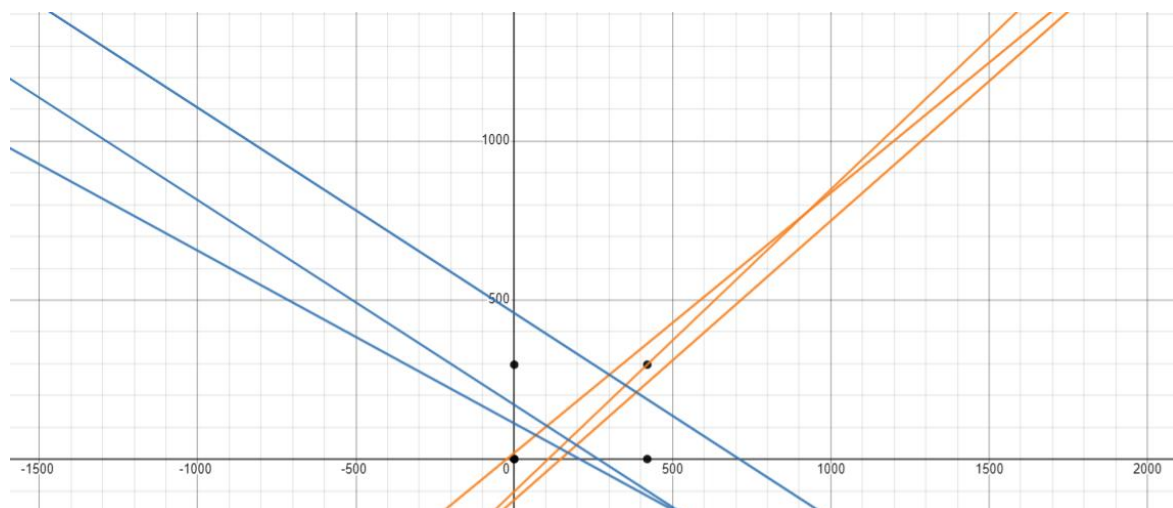


図 6 学習者 A の 1 回目の紙箱ドローイングの可視化パース（上）と
指導者の 2 回目の紙箱ドローイングの可視化パース（下）

は、消失点を意識したアウトライン描画の評価に繋がると考えられる。次に、消失点構成辺の交点同士の距離の総和は、消失点構成辺が 1 つの消失点に向かっていることを前提として、それがどのくらいの精度かを知ることができる。つまり、距離の総和が少ないほど、消失点に向かってストロークを引く技術が高いといえる。よって消失点構成辺の交点同士の距離の総和は、消失点に向かって正確なストロークを引く技術の評価に繋がると考えられる。そして、2 つの消失点の y 座標の距離の差は、2 点透視法において 2 つの消失点がアイレベルからどの程度ずれているかを表している。よって、2 つの消失点とアイレベルとの関係を意識したドローイングの評価に繋がると考えられる。

ここで、交点の数に注目すると、初学者である学習者 A と学習者 B については、3 つの交点を有する

消失点構成辺が存在しなかった。それに対して、ドローイング学習経験を有する学習者 C と指導者については、3 つの交点を持つ消失点構成辺が複数データにおいて確認された。また、3 名の学習者の各ドローイングの交点の数の総和を図 7 に示す。図 7 より、ドローイングの回数を重ねるにつれて交点の数が増加していく傾向が確認できる。なお、学習者 B の 3 回目の紙箱ドローイングについては、1 点透視法で描かれているため、交点の総和が少なくなったと考えられる。

一方で、交点同士の距離の総和と 2 つの消失点の y 軸の距離の差については、3 名の学習者の値の変化幅は、指導者の値の変化幅に比べて大きい傾向にある。このことから、指導者は常に消失点を意識してアウトラインを描いており、消失点に向かってストロークを引く技術も高いと評価することができる。それに対し

表 1:各紙箱ドローイングの解析結果

対象者		交点の数 [個]	交点の距離 の総和[mm]	2つの消失点の y座標の距離の 差[mm]
A	1回目			
	消失点 1	2	1999	-
	消失点 2	0	-	
	2回目			
	消失点 1	2	6550	3097
	消失点 2	2	1244	
	3回目			
	消失点 1	2	503	153
	消失点 2	2	364	
B	1回目			
	消失点 1	0	-	-
	消失点 2	0	-	
	2回目			
	消失点 1	2	2603	11
	消失点 2	2	2057	
	3回目			
C	消失点 1	1	-	1点透視法
	1回目			
	消失点 1	2	1389	801
	消失点 2	2	6015	
	2回目			
	消失点 1	2	3751	833
	消失点 2	3	187	
	3回目			
	消失点 1	3	823	980
	消失点 2	2	4749	
指導者	1回目			
	消失点 1	3	2344	49
	消失点 2	3	1764	
	2回目			
	消失点 1	3	574	76
	消失点 2	3	1507	
	3回目			
	消失点 1	3	695	34
	消失点 2	3	1578	

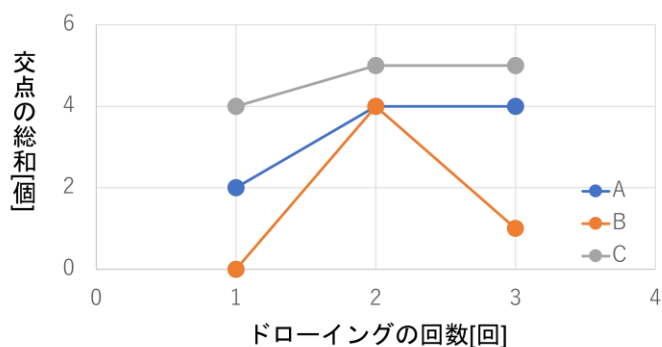


図 7 学習者のドローイング回数と交点の総和の関係

て学習者は、消失点を意識できていなかったり、ストロークを引く技術が未熟であったりすることが定量評価項目の数値から評価できる可能性がある。また、パースが習熟し、消失点構成辺が 3 つの交点を持つようになると、数値による他者との比較が可能となり、評価項目の数値の解釈が多様になると考えられる。しかし、この交点同士の距離の総和と 2 つの消失点の y 軸の距離の差に関しては、交点の数に依存する数値となるため、厳密な比較方法は今後の検討課題となる。

これらのことから、表 1 に示した解析結果は定量的な成長指標として適用できる可能性があると考えられる。特に、交点の数は「消失点を意識して描けているか」というパースを習熟していく上で重要な要素を評価できる可能性がある。

4. スキル獲得支援への適用例

本章では、前章までに示した解析手法と結果の解釈の成果に基づくドローイングスキル獲得支援機能として、パースの可視化と直線精度評価を提案する。

4.1 パースの可視化による評価機能

先行研究である遠隔ドローイング学習支援システムに対して、学習者のドローイングプロセスデータを基に図 5 や図 6 のような形でパースを可視化する機能を導入する。これにより、学習者は自身の描いたストロークが消失点に収束するか否かを視覚的に確認することができる。さらに、3.3 の定量評価項目の値を表示することで、どの程度消失点を意識できているかをフィードバックすることができる。さらに、消失点構成辺の交点が 3 つに満たなかった場合に、どの

辺をどの方向に描けば交点が 3 つになるかをナビゲートできる機能も有用であると考えられる。

4.2 直線の精度評価機能

ドローイングでは、短い線を重ねてアウトラインを表現する方法が存在する。一方、中上級者のドローイングにおいては、長い 1 本のストロークを思い通りの線として描けるか否かが重要となる (6)。先行研究により、長いストロークで円を描くための支援システムが提案されている (6)。一方で、長い直線を描くための支援機能は未だ提案されていない。

それに対して、本稿で述べたストロークの座標データに線形回帰をおこなう手法を適用することで、直線の精度を評価する機能が実現できる可能性がある。学習者が描いた長い直線の座標データに対して、線形回帰の決定係数を求める。この決定係数を評価指標として、描かれた直線と一次方程式の直線との近似の程度を定量的に評価する。この評価をリアルタイムでフィードバックすることで、学習者は直線を上手く描いた際の身体動作を自覚しやすくなると考えられる。直線の精度を評価する機能は、円を描くことに対する支援機能よりも初歩的なものであろう。初学者が紙箱などの単純なモチーフをドローイングする際の支援に役立つことが期待される。

5. おわりに

本研究の目的は、ドローイングにおけるパースの習熟度を定量的に評価するための指標の具体化である。本稿ではまず、遠隔ドローイング学習支援システムと、それによって蓄積されたプロセスデータについて説明した。次に、パースの定義を示し、ドローイングプロセスデータに対する線形回帰による解析手法と評価指標を提案した。そして、複数の学習者と指導者のドローイングに対する解析結果を比較することで、パースに関する成長指標としての可能性を考察した。

今後は、本稿で提案した解析手法と評価指標の妥当性をさらに探究する。そのために、今回は解析対象としなかったドローイングプロセスデータを解析していく。そして、提案指標のスキル評価性能を精査した上で、ドローイング支援機能として実装していく。

謝辞

遠隔ドローイング支援システムを 2012 年より美術教育に導入頂き、多くの助言を頂いている日本外国語専門学校留学科海外芸術大学留学コースの教員と学生の皆様に感謝いたします。本研究は科研費 23K17015 により支援されています。

参考文献

- (1) 佐藤聖徳, “美術・デザイン系大学におけるデッサン指導の発展的試み”, 静岡文化芸術大学研究紀要, 4, pp.153-162, 2004.
- (2) 関根英二, “美術体系の試み”, 美術教育学会大学美術教科教育研究会報告, 6, pp.89-100, 1984.
- (3) Takashi Nagai, Mizue Kayama, Kazunori Itoh, “A Drawing Learning Support System based on the Drawing Process Mode” Interactive Technology and Smart Education, 11(2):146-164 2014(Apr.)
- (4) 遠近法-Wikipedia
<https://ja.wikipedia.org/wiki/%E9%81%A0%E8%BF%91%E6%B3%95> (2023 年 6 月 11 日確認)
- (5) 浅井雄大, 香山瑞恵, 館伸幸, 永井孝, “ドローイングプロセスのアウトライン描画段階における定量的特徴に関する基礎的検討”, 教育システム情報学会北信越支部学生研究発表会, 2023
- (6) 山本景子, 安田和夫, 山本豪志朗, 倉本到, 辻野嘉宏, “学習者自身の運筆動作で教示する個性を伸ばすためのドローイング学習支援システム”, 情報処理学会インタラクシオン 2014 論文集, pp.554-561, 2014
- (7) Pengli Du, Yingbin Liu, “The Techniques and Evaluation Method for Beautification of Handwriting Chinese Characters Based on Cubic Bézier Curve and Convolutional Neural Network”, International Conference on Computer Science & Education (ICCSE), 2019

顔検出と教員の発話による

授業中の学習者の集中度評価システムの開発

左座祐之助^{*1}, 中村伊吹^{*1}, 尾崎剛^{*1}, 広瀬啓雄^{*1}

^{*1} 公立諏訪東京理科大学大学院

Development of an Evaluation System for Learners' Concentration Level in Class by Face Detection and Teacher's Speech

Yuunosuke Zouza^{*1}, Ibuki Nakamura^{*1}, Takeshi Ozaki^{*1}, Hiroo Hirose^{*1}

^{*1} Suwa University of Science

Learners' facial orientation was photographed, and the relationship between facial orientation and the teacher's speech content was visualized to assess the learners' level of concentration. The results of the experiment showed that there was a correlation between the teacher's speech content and the learner's face orientation, indicating that the teacher's active speech increased the learners' interest and they tended to turn their faces. In addition, the time for learners to turn their faces includes the time for working on assignments and checking slides, and this information may also be useful for class improvement.

キーワード: 集中度, 学習支援, 顔検出, 文字おこし

1. はじめに

授業中, 教員が学習者の注意を引き続けることは, 学習成果の向上や, モチベーションを維持するうえで必要不可欠である. しかし, 教員が常に注意をひくことは容易ではなく, また, 教室全体の集中度を測るのも困難である.

左座(2022)では, 学習者のモチベーションを把握及び維持・向上させ, 授業改善の支援方法を明らかにすることを目的とした研究が行われた. この研究では, 達成動機理論を用いて学習者のモチベーションの数値化し, ARCS 動機づけモデルを活用した授業改善支援システムの構築及び授業改善を行なった. それぞれ学習者にアンケートに回答してもらい, 結果を教員に提示した. 構築したシステムを教員が利用し, 授業改善をした結果, 学習者のモチベーションが向上したことが確認できた. しかし, 授業を重ねるごとにアンケート

回答者数が減っており, マンネリ化などが問題となり, また, アンケート手法は主観的な結果になってしまうという問題が挙げられた.

また, 田中(2010)では, プロジェクタを使用した講義において, 表示切り替え時における学生の動きや顔の向きを検出し, 相関を調べることで, 学生の授業に対する集中度を客観的に評価するシステムを開発した. また, この手法を板書で行った場合の研究がされた. 結果として, ある程度以上の大きさで写っており, 受講態度がはっきりしている学生に対しては, 目視による判定に対して 70%程度の精度で判定できることが確認された. しかし, プロジェクタを使用した表示切り替えに対して, 板書では検出精度が低い結果となった.

そこで, 本研究ではARCS動機づけモデルの「注意」の側面に焦点を当て, 授業中の学習者の顔を撮影し,

顔の向きから集中度を評価する．また，教員の発話内容を文字おこしし，発話した文字数と学習者の顔の向きとの関係を可視化することを目的とする．

2. システム概要

2.1 RetinaFace

RetinaFace は，顔検出や顔の特徴点検出のための深層学習ベースのアルゴリズムである．顔の位置を検出するだけでなく，顔の主要な特徴点（目，鼻，口）を検出することも可能である．また，小さな顔やスケールの異なる顔，マスク着用している場合でも検出できる．

2.2 顔の角度

RetinaFace で検出された顔の特徴点の座標をもとに顔の向きを計算する．顔の向きによって，目，鼻，口の位置が変わる．それを利用し，特徴点間の距離を求め，そこから角度を求めるようにした．図 1 は，右目・左目・鼻の特徴点から角度を求める例である．同様に，右目・鼻・右口，左目・鼻・左口，鼻・右口・左口の 3 か所の角度も求める．また，本研究では正面を向いている学習者を検出するため，顔の角度に閾値を設定し，顔の向きを判別することにした．

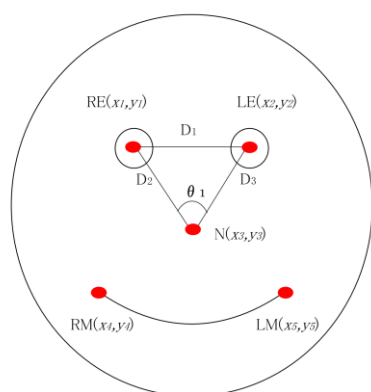


図 1 顔の向き概要

2.3 文字おこし

文字おこしは，Python の音声認識ライブラリである SpeechRecognition を使用し，Google の音声認識 API を利用した．本研究では，授業中に教員が話している内容を録音しておく．その後，SpeechRecognition ライブラリを使用し，1 分間隔で音声を取り出し，Google の音声認識 API を利用し，音声をテキストに変換し

csv に保存した．

3. 実験

3.1 実験方法

3.1.1 分析対象科目

本実験を行うにあたり，公立諏訪東京理科大学工学部情報応用工学科で開講している「データベース論及び演習」を分析対象とした．

「データベース論及び演習」は，3 年次選択科目であり，100 分×2 コマで 1 回分の授業となっている．この授業では，IoT などから収集されるビッグデータを，データベースに格納し，AI や機械学習および統計的手法で分析するために，必要なデータを検索する．さらに，分析可能な状態の前処理の基礎を学習する．また，SQL, PHP, Python を用いて授業を行っている．

3.1.2 実施方法

対象科目の授業中に，教室全体が映るように，カメラをモニターと同じ高さに配置した（図 2）．カメラには広角で教室全体が撮影できる Gopro10 を利用した（図 3）．また，撮影中に教員の音声のみを録音できるように，教員にトランスミッター（図 4 左）をつけてもらい，Gopro10 にチューナー（図 4 右）を接続した．



図 2 教室撮影風景



図 3 Gopro10



図 4 トランスミッターとチューナー

3.2 実験結果

3.2.1 分析間隔 1 分間ごと

「データベース論及び演習」の第 2 回で、1 分間隔で顔の向きは考慮せずに検出を行った。結果は図 5 のようである。横軸は時間（分）を表しており、縦軸は文字数と顔の数を表している。また、棒グラフが 1 分間ごとの教員の発話した文字数であり、赤色の折れ線グラフが 1 分ごとに検出した顔の数、緑色の折れ線グラフが移動平均である。図 5 より、顔を検出しただけでは、顔の数に変化が見られなかった。1 分間ごと検出しているため、検出のタイミングで下を向いており、検出されていない可能性も考えられる。

そこで、次に 10 秒間隔で検出を行った場合と、顔の向きを考慮した検出を行った。

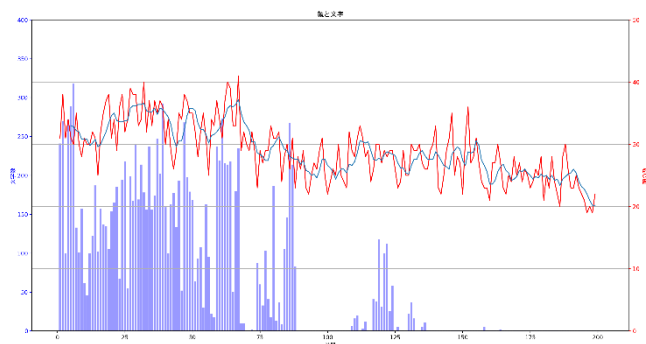


図 5 1 分間ごとの顔検出推移

3.2.2 分析間隔 10 秒ごと

10 秒間隔で検出を行った結果は図 6 のようになっている。縦軸、横軸は図 5 と同様であり、赤色の折れ線グラフが 10 秒ごとに検出した顔の数である。

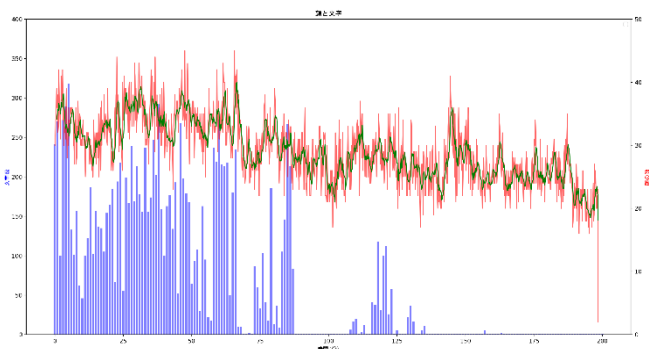


図 6 10 秒ごとの顔検出推移

次に 2.2 の顔の向きを考慮した結果は図 7 のようになっている。教員の発話数が多いと正面を向いている学生が多くなり、発話数が少ないと検出した顔数も減少した。第 3 回での結果は図 8 のようになっており、第 2 回と同様、発話数によって、顔の数に変化していることが分かる。

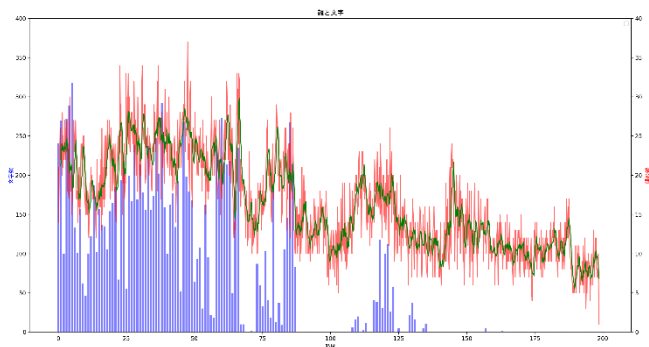


図 7 10 秒ごとに顔の向きを考慮した検出推移
(第 2 回)

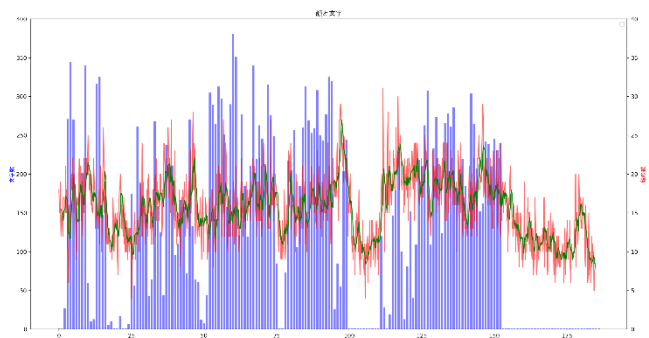


図 8 10 秒ごとに顔の向きを考慮した検出推移
(第 3 回)

4. 考察

4.1 顔と文字数の相関

検出した顔の数と教員の発話した文字数の相関を調べた。なお、顔の数は顔の向きを考慮した結果を用いることとし、教員の発話した文字数とデータの数と同

じにするため、10 秒ごとのデータを 6 回分の移動平均を算出した。第 2 回の授業での結果は図 9 のようになっている。相関係数は 0.752 と正の強い相関があった。また、第 3 回でも 0.481 と正の相関があり（図 10）、教員が積極的に発話することにより、学習者の関心が高まり、顔を向ける傾向がある可能性が考えられる。しかし、第 2 回に比べて第 3 回は相関係数が低下している。この原因として、第 3 回では教員の発話数が多く、関心がそれてしまった可能性が考えられる。授業に抑揚を持たせるなどを行うことで、関心を維持できると考える。また、第 3 回目以降で相関係数が低下していれば、授業のマンネリ化なども考えられるため、引き続き調査していこうと考える。

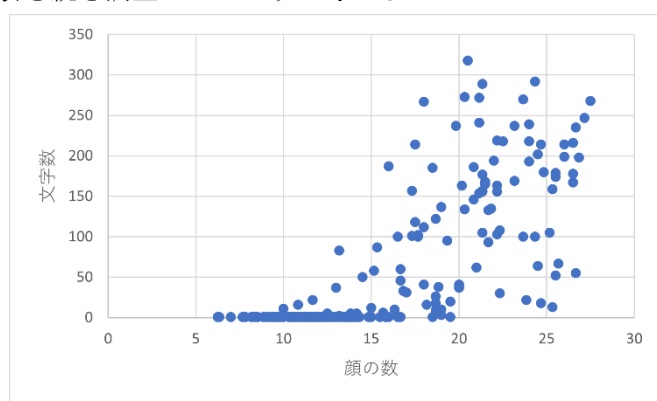


図 9 発話文字数と顔の数の相関（第 2 回）

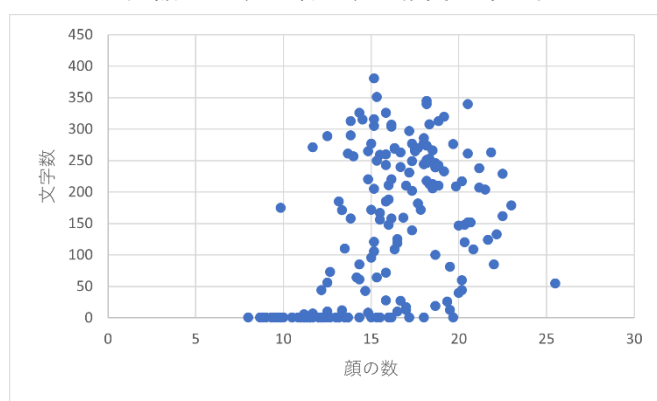


図 10 発話文字数と顔の数の相関（第 3 回）

4.2 話してないのに顔の数が多いとき

図 7、図 8 において、教員が発話していないが、発話時と同数顔を検出している。これは課題に取り組んでいる時間に、わからない箇所を正面のモニターでスライドや学習した内容を確認しているからだと考えられる。学習者の顔だけでなく、スライドも撮影することで、学習者がどこで苦戦しているかなども把握することができ、授業改善に役立つと考えられる。

4.3 学習者の顔検出率

第 2 回、第 3 回での学習者の出席者数は表 1 のようになっている。出席者数に対して顔の検出率は、第 2 回で 39%、第 3 回で 35% が最大であった。PC を見たり、別の方向を見ている学習者もいるが、前の人に顔が隠れていることや教室の奥に座っている学習者の検出ができていないことが検出率の低さの 1 つの要因だと考える。カメラの位置を変更することや、カメラを複数設置することで検出率が上がるのではないかと考える。また、本研究では約 120 名を対象にしたため広い教室での調査となったが、一回り小さい教室で対象が少なければ、より検出精度が上がると思う。

表 1 授業の出席者数

授業回数	出席者数（人）
第 2 回	115
第 3 回	115

5. まとめ

本研究では、ARCS 動機づけモデルの「注意」の側面に焦点を当て、授業中の学習者の顔を撮影し、顔の向きと教員の発話数の関係を可視化することを目的とした。

授業中に教室全体が映るように、カメラをモニターと同じ高さに配置し、撮影中に教員の音声のみを録音できるようマイクを設置した。

顔の検出には RetinaFace という、顔検出や顔の特徴点検出のための深層学習ベースのアルゴリズムを使用し、得られた特徴点から顔の向き（角度）を計算し、正面を向いているかどうか判別するようにした。

文字おこしは、Google の音声認識 API を利用し、音声をテキストに変換し csv に保存した。

10 秒間隔で顔の向きを計算し、文字おこしは 1 分間隔で行うようにした。その結果、発話数によって、顔の数が増えていることがわかった。また、検出した顔の数と発話した文字数には正の相関があり、教員が積極的に発話することにより、学習者の関心が高まり、顔を向ける傾向があることが確認できた。

今後の課題として、本研究では教員の発話した文字数と顔の向きから集中度を測定したが、発話した内容

を分析し、どのような発言に学習者が興味を示すのかを調査していきたいと考えている。また、学習者が座っている位置によって顔の向きも変わってくるため、座席によって閾値を変化させていこうと考える。

謝辞

本研究は JSPS 科研費 22K02893, 23K02715 の助成を受けたものです。

参 考 文 献

- (1) RetinaFace | InsightFace: an open source 2D&3D deep face analysis library,
<https://insightface.ai/retinaface> (2023 年 6 月 1 日確認)
- (2) GitHub-UberI/speech_recognition: Speech recognition module for Python, supporting several engines and APIs, online and offline,
https://github.com/UberI/speech_recognition (2023 年 6 月 1 日確認)
- (3) 左座祐之助, 広瀬啓雄, 尾崎剛: “学習者と教員の授業に関するモチベーションを可視化するシステム構築及びその効果”, 第 84 回全国大会講演論文集 2022(1), pp.135-136 (2022)
- (4) 尾形直哉, CHAKRABORTY Goutam, 馬淵浩司, 松原雅文: “角度の変化に対応した顔認識手法”, 情報科学技術フォーラム, FIT 2007 号, 一般講演論文集第 3 分冊, pp.61-62 (2007)
- (5) 田中敏光: “学生の動きと教示情報との相関から講義中の学生の集中度を評価するシステムの開発”, 科研費 20500814, 2010 年度研究成果報告書 (2010)

呼吸活動での胸郭運動計測の試み

伊藤敏^{*1}, 井上祥史^{*2} 鷲野嘉映^{*3}

^{*1} 岐阜聖徳学園大学 ^{*2} 岩手大学 ^{*3} 愛知みずほ短期大学

Attempt to measure thoracic cage movement in respiratory activity

Itou Satoshi^{*1}, Inoue Shoshi^{*2} and Washino Kaei^{*3}

^{*1} Gifu Shotoku Gakuen University, ^{*2} Iwate University ^{*3} Aichi Mizuho Junior College

The proportion of elderly people in the total population is increasing in Japan, and maintaining and promoting the health of the elderly is an important issue. The elderly person are characterized by sympathetic dominance in autonomic nervous system responses. Breathing exercises have been reported to be useful in regulating the autonomic nervous system. In this study, we propose methods that use depth information from a depth camera and the changes of flexion sensors as a way to understand the areas used for respiratory activity.

キーワード: 胸郭運動の計測, 呼吸, 曲げセンサ, 深度カメラ

1. はじめに

我が国の高齢者の総人口に占める割合は 2015 年で 26.7%を超えており⁽¹⁾, 平均寿命は女性で 87 歳, 男性で 81 歳を超え(2021)⁽²⁾, 高齢者の健康保持増進は重要な課題である. 高齢者の特徴として, 自律神経系の反応においては交感神経が優位である⁽³⁾. 自律神経系の調整には呼吸法が有用であることが報告されている⁽⁴⁾. 高齢者の呼吸法においては, 胸式呼吸より腹式呼吸が効率的である⁽⁵⁾. 腹式呼吸は, 横隔膜を意識的に大きく動かす呼吸である. 横隔膜は肺の下にある筋肉で, その周辺には神経組織が集まっているため, 横隔膜の動きを介して副交感神経が刺激を受けることで自律神経が調整される⁽⁶⁾. 高齢者の意識的腹式呼吸中には副交感神経が優位になり, リラクゼーション効果をもたらす, ストレスホルモンの分泌が抑制されることが報告されている⁽³⁾. よって, 高齢者においては, 自らの呼吸の状態を把握し, 日常的に腹式呼吸を意識的に行うことが重要であると考ええる.

本報告では, 胸郭部のどの部分を利用して呼吸活動をしているかを客観的に把握する方法をいくつか提案

する. 呼吸を簡易的に非侵襲的に把握する方法として, 曲げセンサおよび depth カメラの深度情報から胸郭部の変動を検出する方法について比較検討した.

本稿では, 2 章で呼吸検出の方法として曲げセンサおよび depth カメラからの深度情報より胸郭の動きを推定する方法を記述し, 3 章で適応例として, 立位での呼吸行動検出を試み, 4 章でまとめる.

2. 呼吸検出方法

人は呼吸をするために, 外肋間筋と横隔膜を収縮・弛緩させている. これらの収縮により胸郭や腹腔は前にせり出し, 吸気時には元に戻る. すなわち, 呼吸活動により胸郭や腹部に物理的動きが生じる.

呼吸活動に伴う胸郭の変動について考察する. 一回換気量 (tidal volume) とは一回の呼吸運動 (呼気と吸気) で気道・肺に出入りするガスの量のことを指す. 単位は mL で表現している. 成人の安静時における一回換気量は「理想体重×6[mL/kg]」で算出する. 理想体重は次式で計算可能である⁽⁷⁾.

男性[kg] = $50 + 0.91 \times (\text{身長[cm]} - 152.4)$

$$\text{女性[kg]} = 45.5 + 0.91 \times (\text{身長[cm]} - 152.4)$$

これらの式より，一回換気量は 250 から 450mL 程度である．胸郭内部の体積は肺活量程度と考えると 4000mL から 5000mL 程度と推定される．呼吸に寄与する胸郭部を円筒と仮定する．さらに，円筒の高さを変動しないと仮定すると，呼吸に伴う変動は円筒の半径の変動として観察される．

一回換気量は胸郭体積の 10%程度と推定される．故に半径の変動は 3%程度と思われる．胸囲が 90cm 程度と仮定すると，曲率半径の変動は $\pm 0.4\text{cm}$ 程度となる．

この呼吸に伴う胸郭部の膨張・収縮の動きを，計測するには直接胸郭部に接触させて変動を計る方法と胸郭部の変動を非接触でとらえる方法に大別される．本報告では曲げセンサによる方法を提案する．非接触では depth カメラによる深度情報からの方法を提案する．一連の動きは擬周期的に変動すると考えられる．これらの方法を用いて呼吸状態を同時計測して両方法の違いを検討する．なお，今回は呼気吸気量の計測および推定は行わない．

2.1 曲げセンサによる方法

曲げセンサは，一方向の曲げ程度により抵抗値が変化する Bending Detection Sensor (TAIWAN ALPHA ELECTRONIC CO., LTD 製)を用いた⁽⁸⁾．用いた曲げセンサの抵抗値はフラットなとき 25k Ω 程度，強く曲げられたとき 180k Ω 程度に変化する．

図 1 に示すように，曲げセンサの両端にフラットな板を貼り，「曲げ」が「曲げセンサ」の一部だけに集中する構造にする．曲率半径の変化を曲げセンサの角度変化として計測する．曲げセンサの抵抗値検出は抵抗による分圧法で行い，曲率半径が増加する場合，すなわち胸郭部が膨張（吸気時に相当）する場合に計測値が上昇する様に設定をした．

作成した装置を図 2 に示す．センサから得られたデータをマイクロプロセッサで処理をして SD カードへ時間とともに記録した．サンプリング周波数は 50Hz とした．装置はセンサ部のサイズ 120 $\times$ 10 $\times$ 2mm，抵抗値検出記録部が 55 $\times$ 55 $\times$ 25mm で構成され，電池で作動する．SD カードに保存されたデータは，雑音除去のため PC で 1.0Hz のローパスフィルタを掛け，解

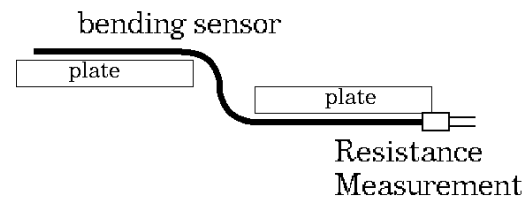


図 1 曲げセンサの形状

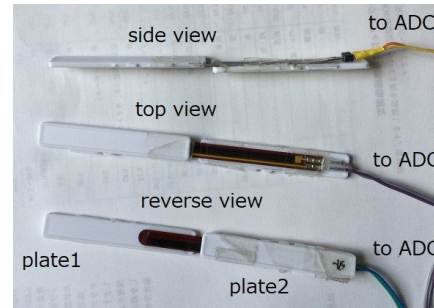


図 2. 曲げセンサ装置の写真

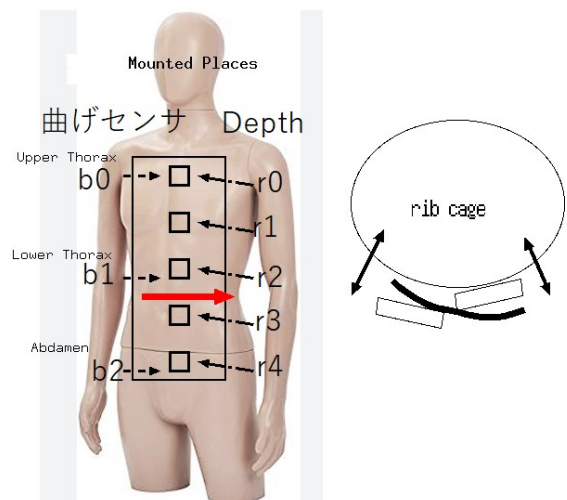


図 3 装置装着場所 b0, b1, b2 と曲げセンサ配置
および深度情報取得場所 r0 ~r4
慣性センサの y 軸方向（赤矢印）

析に供された．

曲げセンサを胸郭部に配置した様子を図 3 に示す．胸郭上部（第 3 肋骨付近）b0，胸郭下部（第 10 肋骨付近）b1，腹部下部（へそのやや下）b2 の 3 カ所に配置し，それぞれを b0, b1, b2 と名付けた．

呼吸検出に実績を持つ慣性センサを胸郭下部に配置し，同時計測を行うことで検証をした．曲げセンサおよび慣性センサは加圧シャツに縫付けて実験に用いた．

2.2 深度カメラによる方法

呼吸に伴う胸部部の膨張・収縮の動きを、胸部部を撮影した深度カメラの深度情報の変化として記録する。一連の動きは擬周期的に変動すると考えられる。

深度カメラは Intel RealSense Depth Camera D415 を用いた。Depth 計測の画角は 70° である。D415 カメラは中央に RGB カメラ、その両脇に depth カメラを持ち視差から深度を求める方式である（図 4）。各 pixel の RGB カメラからの色情報と depth カメラからの深度情報の画角合わせはソフトウェアで行った。深度情報取得処理などは Intel Realsense SDK を用いた⁹⁾。なお、深度情報の精度は 2%とされている。

撮影は図 4 に示すような配置で、距離約 800mm で撮影領域に胸部部を含み、15fps、解像度 424×240 で行った。深度情報から体の肋骨上部から下腹部までに該当する領域、横幅 100pixels 縦 150pixels、を取り出した（図 4 の右図）。これを”解析領域”とする。解析領域の中央線上に図 3 で示した 5 領域の値を抽出し、 r_0 , r_1 , r_2 , r_3 , r_4 とした。5 領域は、それぞれ胸部最上部 r_0 、上中部 r_1 、胸部中央部 r_2 、胸部下部 r_3 、腹部下部 r_4 に相当する。抽出した 5 領域は、それぞれ 11×11 pixels の 121pixels の平均値とし、解析領域(100×150)の最大値 $d_{\max}$ と最小値 $d_{\min}$ を基準に、次式より 0 から 1 の間に d_{normal} として規格化した。

$$d_{\text{normal}} = (d_{\max} - d) / (d_{\max} - d_{\min})$$

d : 実験より得た値

この規格化により、実験中の体の前後の揺れがもたらす変動を補正し、かつ胸部部が膨張（吸気時に相当）したときに値が増え、縮小（呼気時に相当）したときに値が減る様にした。

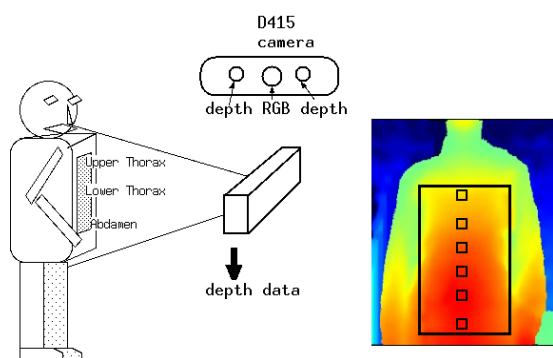


図 4. Depth カメラ撮影範囲（左）と深度イメージ（右）

実験は Realsense SDK の録画機能を用いて ROS 形式の bag ファイルとして保存され、その bag ファイルより取り出されたデータは、雑音除去のため PC で 0.5Hz のローパスフィルタを掛け、解析に供された。

実験時には上半身に密着したシャツを着て行った。また曲げセンサとの同時計測時にはアンダーシャツの上にワイシャツを着て、その上から加圧シャツを着て行った。

3. 適応例 評価実験

提案手法で呼吸検出を評価した。曲げセンサ装置を加圧シャツに縫い付けて上からベルトで体前面に装着し計測を行った。同時に depth カメラで体の前面より録画した。呼吸計測が可能であるかの検証に慣性センサ⁽¹⁰⁾と同時計測し比較検討をした。慣性センサは胸部下部の曲げセンサ b_1 の位置（図 3 参照）に慣性センサ

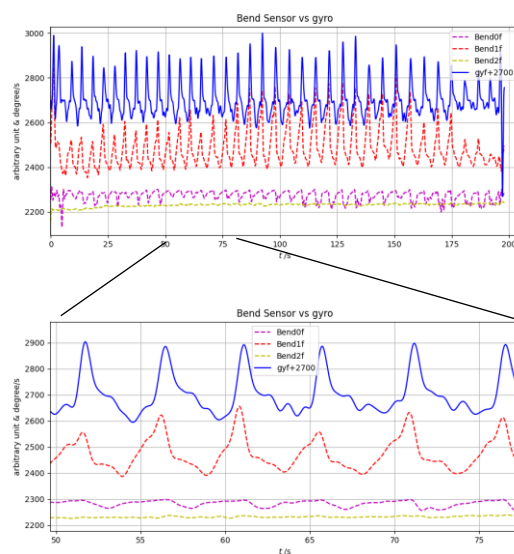


図 5. 腹式呼吸 曲げセンサ（破線）と y 軸周りの角速度（青の実線）下図は拡大図

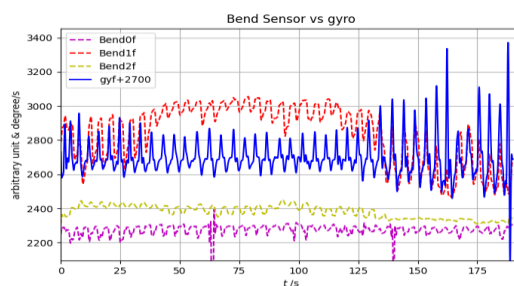


図 6. 胸式呼吸 曲げセンサ（破線）と y 軸周りの角速度（青の実線）

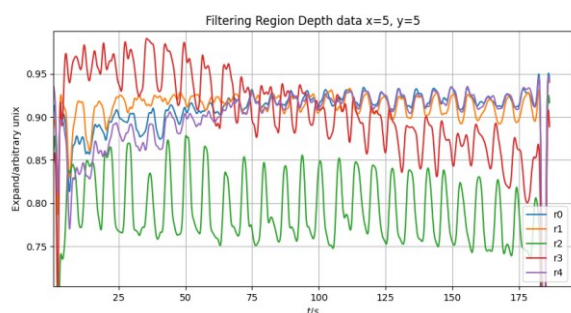


図 7. 腹式呼吸 depth カメラによる胸部各部位の深度

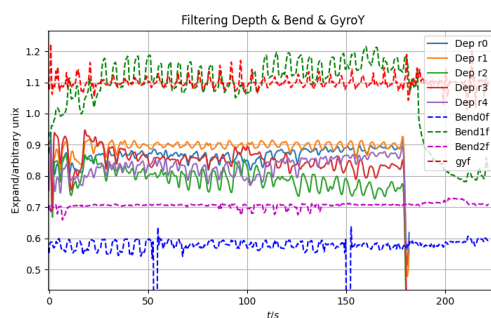


図 9. 胸式呼吸 深度と曲げセンサと角速度

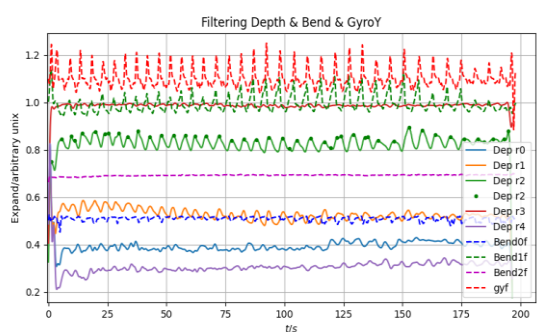


図 8. 腹式呼吸 深度（実線）と曲げセンサ（破線）と角速度（破線赤）

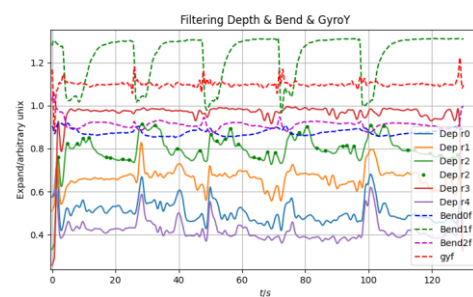


図 10. 深呼吸 5 回
角速度(赤破線), 曲げセンサ(破線), 深度(実線)

サ y 軸正方向が水平で体表面に密着し、左へ向かう方法へ装着した(図 3 の赤矢印)。

立位での腹式呼吸時の曲げセンサと慣性センサからの角速度の時間変化を図 5 に示す。左右の肋骨を結んだ線に平行な軸の角速度は呼吸を表すことが解っている⁽¹⁰⁾。角速度と胸部下部に装着した曲げセンサの変動は完全に同期しており、曲げセンサは呼吸に伴う胸部部の変動を検出できている事を示す。胸部上部に着けた曲げセンサ b0 の変動幅は小さいものの胸部下部 b1 の動きと一致している。一方で下腹部に着けた曲げセンサ b2 の変動は解析に供することが出来ないくらい小さい。

図 6 に立位での胸式呼吸をした場合を示す。3 カ所に着けられた曲げセンサの変動は胸式呼吸と腹式呼吸では変動の仕方が変化し、腹部中央の b1 の変動幅は小さくなり胸部上部の b0 の変動幅が増す。これは胸式呼吸、腹式呼吸を本方法で区別可能であることを示唆している。

立位での腹式呼吸時の胸部部位の深度カメラによる時間変化を図 7 に示す。胸部の全領域で呼吸に伴う変

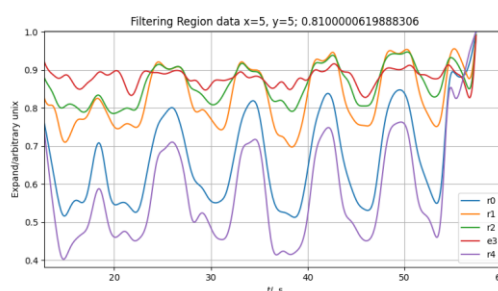


図 11. 深呼吸 5 回 深度（実線）

動が見られる。実験では呼吸回数 26 回を行ったが、領域 r1（橙色実線），r2（緑色実線）では明確に 26 回の繰り返しが見られ、深度情報から呼吸に伴う胸部の変動をととしてとらえることが可能であることを示している。図 8 に深度と共に曲げセンサ、慣性センサからの角速度を同時計測した結果をグラフ化したものを示す。それぞれの変動が一致している。胸式呼吸をした場合にも同様に呼吸回数 32 回に対し 32 の変動の繰り返しが見られた。図 9 に胸式呼吸の場合の図を示す。変動領域が胸部上部で見られる。

深呼吸を 5 回繰り返した場合の深度と曲げセンサと角速度の変化を図 10 に示す。変化が一番大きいのは

胸郭下部であり、吸気時の行動として胸を膨らませている。それに伴い胸郭下部が一端へこみ現象が見られる。他の実験協力者の深度情報から推定した5回深呼吸の様子を図11に示す。この計測ではアンダーシャツのみを着て実施した。胸郭部の複数箇所での変動を計測することで胸郭部の使い方が異なることが解る。このように呼吸に伴う胸郭部の使い方を観測することが出来る。

4. まとめ

曲げセンサによる胸郭部の拡張収縮計測を通じて呼吸を検出することができた。検出する物理量は抵抗値であり、装置が簡便である。曲げセンサの装着はバンドを体につけるだけなので簡単である。曲げセンサは胸郭の膨張収縮のみで変化するため、「曲げ」のみの検出で呼吸を観測する本方法は、行動中の呼吸検出に有効な方法と思われる。また、今回は立位での計測を行ったが姿勢によらずに計測が可能である。伏臥位を除いて就寝中での計測も可能である。曲げセンサの装着場所は今回の計測から変動が小さかった下腹部(b2)よりももっと上部が好ましい。

深度情報からは胸郭部の拡張収縮をカメラからの深度計測を通じて呼吸を検出することができた。体に装置を付ける必要がないため、実験協力者の負担は軽い。一方で深度の検出精度が2%程度の制約があり、これから最も好ましい測定条件は上半身着衣なしである。これは実験協力者の負担になる。胸郭前面がカメラで撮影可能ならば姿勢に制限はない。

高齢者の呼吸に関係する健康法として、ヨガ、マインドストーム、太極拳などの有効性が提案されている。今後、これらの提案行動と組み合わせ、高齢者が自身の呼吸状態の把握にどのようににつなげるか検討を行う。

謝辞

本研究の一部は科研費(19K03178, 20K03164)の助成を受けた。

参 考 文 献

- (1) 総務省統計局 人口推計(2021年(令和3年)10月1日現在) 結果の要約

<https://www.stat.go.jp/data/jinsui/2021np/index.html>
(2023/06/01)

- (2) 令和3年簡易生命表の概況

<https://www.mhlw.go.jp/toukei/saikin/hw/life/life21/dl/life21-15.pdf> (2023/06/01)

- (3) 田中美智子, 長坂猛, 矢野智子, 小林敏生, 榊原吉一: “意識的腹式呼吸がもたらす高齢者の自律神経反応及びホルモン変化”, 形態・機能, Vol10, No.1, pp.8-16 (2011).

- (4) 坂木佳壽: “腹式呼吸が自律神経機能に与える影響”, 体力科学, 50, pp.105-118 (2001).

- (5) 川崎卓也, 坂井泰, 柿崎藤泰, 西田直弥: “吸気量の変化が横隔膜変位量と腹部・下部胸郭容積変化に及ぼす影響”, 理学療法学, 39 Suppl. No.2, 0266 (2011).

- (6) 柳奈津子, 小池弘人, 小板橋喜久代: “健康女性に対する呼吸法によるリラックス反応の評価”, 北関東医学, 53, pp.29-35 (2003).

- (7) 半井悦朗, 釘宮豊城, 稲田英一: “呼吸管理に必要な呼吸生理”, 日集中医誌(J Jpn Soc Intensive Care Med.), 15, pp.49-56 (2008).

- (8) Bending Detection Sensor 60mm Effective length, TAIWAN ALPHA ELECTRONIC CO.,LTD, MB060-N-221-A02 <http://www.taiwanalpha.com/en/products/25> (20230609 確認)

- (9) Realsense SDK
<https://github.com/IntelRealSense/librealsense>
(20230609 確認)

- (10) Satoshi Itou, Shoshi Inoue Kaie Washino: “Respiration detection using Inertial Measurement Unit”, The 2019 International Symposium on Electrical and Electronics Engineering (ISEE 2019) CES-3:Communications Room:401 B4, 13:30-13:50 Cot11th, (2019)

情報活用学士力を測定するメトリクス導出のための一検討

小俣 昌樹^{*1}, 石田 雪也^{*2}, 金子 大輔^{*3}, 吉川 雅修^{*1}, 古賀 崇朗^{*4}

^{*1} 山梨大学

^{*2} 公立千歳科学技術大学

^{*3} 北星学園大学

^{*4} 佐賀大学

A Study on Derivation of Metrics to Measure Graduate Attributes on Utilization of ICT

Masaki Omata^{*1}, Yukiya Ishida^{*2}, Daisuke Kaneko^{*3}, Masanobu Yoshikawa^{*1}, Takaaki Koga^{*4}

^{*1} University of Yamanashi

^{*2} Chitose Institute of Science and Technology

^{*3} Hokusei Gakuen University

^{*4} Saga University

The purpose of this study is to construct metrics that quantitatively and objectively measure graduate attributes on utilization of ICT, which are the ICT knowledge and skills that college students should acquire. In order to derive items that constitute the metric, the authors selected various ICT knowledge and skills, categorized them, and systematized them as a tree structure. As a result, the tree structure with a total of 70 items in four categories of online collaboration skills, data utilization skills, ICT system development skills, and ICT ethics was developed. The degrees of importance of the categories and the one-depth items in each category were assessed through paired comparisons by the authors. The results of the comparisons among the categories revealed that the importance of data utilization skills, ICT ethics, online collaboration skills, and ICT system development skills, in that order, were found to be the most important.

キーワード: 情報活用学士力, メトリック, 大学情報教育, ICT

1. はじめに

本研究では、大学生が身に付けるべき情報通信技術 (ICT) に関する知識およびスキルを「情報活用学士力」と定義し、それを定量的・客観的に測定するメトリクスを導出することを目的とする。ここでのメトリクスとは、「情報活用学士力」を下位概念化して細目化した基準や指標のことであり、学生が「はい／いいえ」や数段階の選択肢などを選択して回答することを想定している。このメトリクスを学生が把握して自身を定量的に評価することで、ICT に関して大学生が必要とする「情報活用学士力」を確認したり、その中で何を

どのように身につけていくべきかを計画したり、自身の特性を定量的・視覚的に確認したりすることができるようになると考える。したがって、本研究で想定しているこのメトリクスの使用は、授業での試験やテストのように教員が学生を評価したり成績として扱ったりするのではなく、学生自身が継続的にチェックし続けて、自身の状況を随時把握し続けることである。その際、すべてのメトリクスの達成を目指すことは重要ではあるが、それ以上に、学生が学生自身で ICT に関する自身の特性を可視化して把握することがより重要であると考えられる。

この研究の背景は、大学生の入学時・在学時・卒業

時において、本提案のように ICT に関する知識やスキルを定量的・客観的・汎用的に評価するメトリクスがまだ整備されていないことにある。あわせて、著者らのこれまでの研究⁽¹⁾を含め、大学の情報リテラシー教育においては、未だ、OS やオフィススイートなどの操作方法やそれによる文書作成やプレゼンテーション、および情報モラルに焦点が当てられており⁽²⁾、それらの応用となる情報活用にまでは至っていないことにある。

本稿では、メトリクスを構成する項目を導出するため、著者らによるブレインストーミングによって、情報活用学士力として必要と考える、ICT に関するさまざまな知識とスキルを選び出し、分類したりカテゴリ化したりして、木構造で体系化した。さらに、このカテゴリ同士および各カテゴリの深さ 1 の項目同士の重要度について、著者らによる一対比較で評価・分析して尺度図に示した。本稿の研究結果は、以下である。

- ・ 導出の結果、「オンラインコラボレーション力」、「データ利活用力」、「情報システム開発力」、「情報倫理力」の 4 つのカテゴリで総計 70 の葉（評価項目）をもつ木構造に体系化した。
- ・ カテゴリ間の重要度の関する一対比較の結果、データ利活用力、情報倫理力、オンラインコラボレーション力、情報システム開発力の順で重要であることがわかった。
- ・ データ活用力の中の深さ 1 の項目の一対比較の結果、「機械判読可能な統計表の作成」と「的確なグラフの選択」が重要であることがわかった。

2. 情報活用学士力の導出

2.1 導出方法

情報活用学士力のメトリクスの導出においては、本稿の著者 5 名が全体的な概念を議論したうえで、この概念についての情報を収集したり調査したりし、その結果に基づいたブレインストーミングによって評価項目を決定した。いずれの著者も専門分野に教育学を含んでいる。情報収集・調査においては、高等学校情報科「情報Ⅰ」（以降、情報Ⅰと記述する。）の学習指導要領⁽³⁾および教員研修用教材⁽⁴⁾、高等学校情報科「情報Ⅱ」（以降、情報Ⅱと記述する。）の教員研修用教材

⁽⁵⁾、そして、日本国内の無作為に抽出した 100 大学の初年次情報科目のシラバスを参照した⁽²⁾。その後のブレインストーミングにおいては、マインドマップソフトウェアを使い、項目の修正・追加・削除・移動などをおこなったり、項目のグループ化およびグループの追加・削除・修正・移動などをおこなったりして、メトリクスの木構造を構成した。これらの作業において、情報Ⅰに含まれる用語や知識および OS やオフィススイートの操作については、大学入学までに十分に習得していると考え、情報活用学士力の項目には含めないことにした。

2.2 導出結果の 4 つのカテゴリ

前節の導出の結果、情報活用学士力をルートとする深さ 1 の葉に相当する、オンラインコラボレーション力、データ利活用力、情報システム開発力、情報倫理力の 4 つのカテゴリ（図 1 参照）と総計 70 の葉を導出した（表 1 から表 4）。それぞれのカテゴリにおいて、カテゴリ名を深さ 1 とすると、深さ 4 までの葉に細分化された木構造で構成される。この木構造の末端となる葉は、学生が、「はい／いいえ」や「できる／できない」または 5 段階程度の数値で回答するチェック項目である。たとえば、表 1 で示すオンラインコラボレーション力のカテゴリの中の大項目「非同期型コミュニケーション」の中の葉となる「電子メールでの送受信」の項目であれば、「電子メールを送受信できる／できない」という評価項目になり、学生は「できる」または「できない」を選択して回答する。以降の節で、各カテゴリについて詳説する。

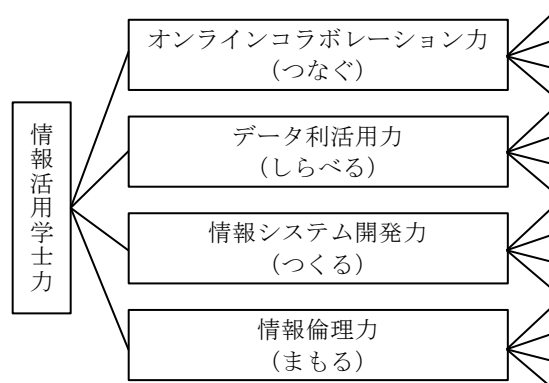


図 1 情報活用学士力の 4 つのカテゴリ

表 1 オンラインコラボレーション力の項目

大項目	小項目	細目
共同作業		
	クラウドサービスでの共同作業	
		ブレインストーミング
		文書の共同執筆
		表計算の共同作業
		共同プレゼンテーション
	メタバースでの共同作業	
	プロジェクトマネジメント	
		自身の作業工程の管理
		他者との作業工程の情報共有
		チームメンバの全員の作業工程の管理
	活動履歴の電子化（e ポートフォリオ）	
		自身の活動の日時・内容・関連ファイルなどを電子的に記録
		自身の活動の電子的な評価／振り返り
		自身の活動を整理して他者へ公開
	電子決済	
同期型コミュニケーション		
	ビデオ会議システムでの通話	
	チャットシステムでの対話	
	メタバースでの対話	
非同期型コミュニケーション		
	電子メールでの送受信	
	Web 上での情報発信	
	アカデミックライティングに基づく電子的な文書作成	

表 2 データ利活用力の項目

大項目	小項目	細目
機械判読可能な統計表の作成		
的確なグラフの選択		
	グラフの特徴を踏まえて、統計データの種類や分析の目的によって使い分ける	
ソフトウェアを使った統計処理		
	尺度（名義尺度，順序尺度，感覚尺度，比率尺度）の違いの説明	
	基本統計量の算出	
	単回帰式の導出	
	検定	
		パラメトリック検定 t 検定の実施 一元配置分散分析の実施
		ノンパラメトリック検定の実施
	因子分析の実施	
	クラスター分析の実施	
プログラミングによる統計処理		
	基本統計量の算出	
	データの視覚化	
		散布図 ヒストグラム 箱ひげ図
	単回帰式の導出	
	検定	
		t 検定の実施 一元配置分散分析の実施

表 3 情報システム開発力の項目

大項目	小項目	細目
開発工程		
	要求定義→設計→実装→テスト	
プログラミング		
	基本三構造	
		順次
		反復
		分岐
	アルゴリズム	
		探索 1 種類以上実装できる 線形探索，二分探索
		単純ソート 1 種類以上実装できる 交換，挿入，選択
	オブジェクト指向	
		概念の説明
	人工知能利用	
		学習器の利用
		学習器の作成
ネットワークング		
	ルータの設定	
		WAN と LAN の接続
		ファイアウォールの設定
	サーバの構築	
		ネットワークへの接続
		ファイアウォールの設定
		http サービスの構築
人間中心設計		
	ユーザビリティ	
		効果・効率・満足度を意識した設計
	情報デザイン	
		場所・辞書順・時間・カテゴリ・階層に注意した情報の整理

表 4 情報倫理力の項目

大項目	小項目	細目
情報セキュリティ		
	マルウェア対策の実施	
	パスワード管理への意識	
ICT 利用に関わる法律		
	著作権法	
		著作物
		著作権
		著作者人格権
	知的財産権	
	個人情報保護法	
	不正アクセス禁止法	
ICT 利用のモラルとリスク		
	発信者のモラル	
		誹謗・中傷・公序良俗に反する投稿をしない
		投稿前に内容を再確認する
		安易な拡散をしない
	受信者のリスク	
		趣味嗜好に合わせてカスタマイズされる
		見たい情報以外がはじかれるフィルターバブルが起きる
		似たような意見や考えが増幅されるエコーチェンバーが生じる
	売買でのリスク	
		サイト運営企業とその中の個々の店舗では信用性に違いがある場合がある
		クレジットカード利用の責任が生じる
		商品の説明や過去の取引評価などの確認が必要である.
		販売側からの直接的な接触の可能性が生じる
	プライバシーのリスク	
		個人のことをどこまでオープンにするか
		つながりと安全を踏まえてアプリやサービスを設定する
		自宅や行動範囲が伝わってしまう可能性がある
ICT の社会への影響		
	ニュースや新聞などで ICT に関する情報に関心をもつ	

2.2.1 オンラインコラボレーション力

オンラインコラボレーション力とは、インターネットなどを介したオンライン上での情報伝達、情報共有、共同作業などに関する知識やスキルである。オンライン授業をはじめ、オンラインのミーティングや研究活動、さらには、卒業後の進路のためのオンライン面談など、さまざまな場面において、インターネットなどを介してつなぐ／つながるための知識とスキルが必要である。さらに、つなぐ／つながるだけでなく、クラウドサービスやメタバースなどでの協調的な作業を通して、共同で文書を執筆したり、データを分析したり、ものづくりに取り組んだりするための知識とスキルが必要である。

表 1 に、オンラインコラボレーション力を細分化した項目を示す。この表の通り、本研究では、いわゆるつなぐ／つながるためだけのための知識やスキルについては、大学入学までに習得していると考え、同期型／非同期型のコミュニケーションの業程度にしか細分化していない。一方で、共同作業については、多くの項目を設けた。特に、インターネットやクラウドを使用した共同作業に細かな項目を設けたり、一回一回の共同作業のみならず、継続的な共同作業のための作業工程の共有や活動履歴の記録のための e ポートフォリオなどの項目を設けたりした。電子決済では、共同作業を進めていくうえでのオンライン決済やメンバー同士でのスマートフォンのアプリケーションを使った金銭授受について問う。

2.2.2 データ利活用力

データ利活用力とは、コンピュータを利用した統計処理、統計分析、検定、およびデータの可視化などに関する知識とスキルである。大学の活動において、ビッグデータをはじめ、機械学習など、膨大な質的データや量的データをコンピュータに処理させる機会はますます増えていく。このための、データの整理方法、基本的な統計の用語、そして、ソフトウェアを使った統計処理に関する知識とスキルが必要である。さらに、今後は学部や学科にかかわらず、Python や R 言語をはじめとするプログラミング言語を使った統計処理のスキルも必要とされていくと考える。

表 2 に、データ利活用力を細分化した項目を示す。最低限必要と考える統計に関する知識について、ICT

とは直接的には関係しないと考えるが、データ利活用には必要と考えて項目に含めた。この理由は、コンピュータを使った統計処理においては、誤ったデータを与えても処理を進めることができしまったり、処理結果の是非を確認せずに利用してしまったりすることに十分な注意が必要と考えたためである。あわせて、可視化も含めた理由は、大学生として、統計処理のみならず、その結果を誤解なく理解してもらう情報提示までを求められると考えたためである。なお、このカテゴリには、ICT ではなくデータサイエンスに含まれる項目も考えられ、その区分が難しいが、データサイエンス分野の方がより適切であると考えられる項目については、本研究のメトリクスには含めないことにした。この表の「機械判読可能な統計表の作成」とは、総務省統計局が発表した「統計表における機械判読可能なデータの表記方法の統一ルール」⁽⁶⁾を指す。

2.2.3 情報システム開発力

情報システム開発力は、プログラミング、ネットワーキング、および開発工程などに関する知識とスキルである。主にソフトウェア開発のための基礎知識となるプログラミングおよびアルゴリズムに関する知識とスキルの項目を設けた。これらに関して、理系／文系または卒業後の進路にかかわらず、今後ますます、さまざまな業務において、簡易なアルゴリズムやそのコーディングに関する知識とスキルを必要とすると考え。さらに、今では必要不可欠なインターネットの利用について、ある程度は自身で構築・設定できることが求められると考える。なお、ハードウェアに関する知識やスキルについては、大学生の「利活用」を中心に考えると、情報 I の範囲で概ね満たしていると考え、ほとんど含めていない。

表 3 に、情報システム開発力を細分化した項目を示す。プログラミングにおいては、使用プログラミング言語を指定せず、どの言語にも概ね共通している基本三構造を評価項目とした。アルゴリズムにおいては、実用的なアルゴリズムに関する項目よりは、アルゴリズムの概念やごく簡易な例を項目とした。人工知能については、利用と作成とに大別した。分野や活動に関わらず、今後は、既に学習済みの学習器を流用して情報システムを開発する活動が考えられるためである。あわせて、情報システム開発では、プログラミングの

イメージが強いと考えられるが、それだけではなく、人間中心のユーザインタフェースを考えたり、情報を整理して提示するためのデザインを考えたりすることも重要であるとする。

2.2.4 情報倫理力

情報倫理力は、使用機器やデータを情報セキュリティインシデントから防ぐための知識とスキル、そして、法律やルールやモラルに基づいて権利などを遵守するための知識である。大学生は、その年齢から、日本の法律ではほとんどが成人となるため、自らの不法行為責任や契約責任がより重くなり、ICTに関する法律においても同様のことがいえる。特に、コンテンツの制作側や発信側として、そしてオンラインでの売買行為において、注意すべきことが増えてくる。さらに、SNSなどにおいて、自身および他者の個人の事柄をどのように扱うのかという倫理観も必要である。

表4に、情報倫理力を細分化した項目を示す。小学校・中学校・高校で学習したICTに関する基本的な権利やモラルについてはあまり細分化せず、大学生として、ひいては成人として、情報の受信者はもとより、発信者としてあるいは売買の当事者として、法律とモラルに基づいた行動に必要な知識やスキルを問う。

3. カテゴリ間および深さ2の項目間の一対比較

前節の情報活用学士力の木構造において、深さ1のカテゴリ同士、そして、各カテゴリの子（木構造全体における深さ2）同士の重要度について、本稿著者5名による一対比較をおこなった。

3.1 比較方法

カテゴリおよび各カテゴリの子の中の2つの項目（一対）のすべての組み合わせを左右に並べたうえで、その間を「左の項目がかなり重要」「左の項目が少し重要」「左右同じくらい重要」「右の項目が少し重要」「右の項目がかなり重要」の5段階に分け、この中の1つを選択した。組み合わせの提示順序およびその左右の並びは、すべての回答者で同じである。

回答者全員が、事前に前節の情報活用学士力の木構造を閲覧してから、すべての一対の重要度を比較・回答した。どの一対においても無回答を許さず、必ず5

段階のうちの1つを選択することとした。なお、いずれの回答者も、この回答以前から一対比較法を十分熟知している。回答の際は、回答者の専門や所属学科などに限らず、全般的な大学生の「学士力」として判断することを指示した。

5名から得た一対比較の回答について、Scheffeの一対比較法による分析をおこなった。一対において、当該項目が「かなり重要」の場合2、「少し重要」の場合1、「左右同じくらい重要」の場合0に（つまり、反対側の項目を-2、-1、0に）数値化した。その後、これらの値をクロス表として集計し、分散分析をおこない、最終的には、重要度に関する比較項目間の距離を表す尺度図で示した。

3.2 比較結果

図2に、4つのカテゴリ間の位置およびヤードスティックを示す。数直線の右側ほどより重要であることを示す。ヤードスティックが0.03であることから、すべてのカテゴリ間において、有意な差（有意水準1%）がある。すなわち、回答者の著者らは、4つのカテゴリの中で「データ利活用力」がもっとも重要であると考えたことがわかる。

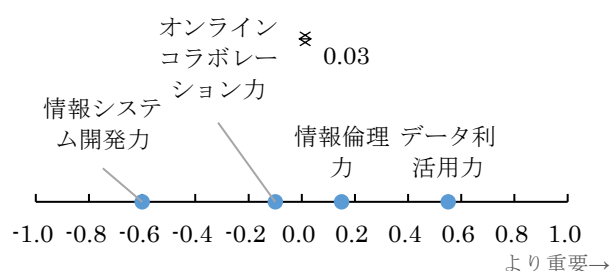


図2 カテゴリ間の重要度の比較結果

図 3, 図 4, 図 5, および図 6 に, 各カテゴリ内での深さ 1 (木構造全体の深さ 2) 同士の各項目の位置およびヤードスティックを示す。いずれも, 数直線の右側ほどより重要であることを示す。図 3 のオンラインコラボレーション力においては, 残差の値が負となったため, ヤードスティックは示していない。したがって, 図 3 の項目間においては, 有意な差はなく, 回答者間でのばらつきが大きかったことがわかる。図 4 およびこのヤードスティックが 0.33 であることから, 「的確なグラフの選択」と「ソフトウェアを使った統計処理」との間, および「ソフトウェアを使った統計処理」と「プログラミングによる統計処理」との間に有意な差 (有意水準 1%) があり, 「ソフトウェアを使った統計処理」と「プログラミングによる統計処理」よりも, 「機械判読可能な統計表の作成」および「的確なグラフの選択」が重要であることがわかる。図 5 およびこのヤードスティックが 0.27 であることから, 「開発工程」と「プログラミング」との間, および「人間中心設計」と「プログラミング」との間に有意差があり (有意水準 1%), 4 項目間に大きな差はないものの, 「プログラミング」がもっとも重要であることがわかる。図 6 およびこのヤードスティックが 0.25 であることから, 「ICT 利用に関わる法律」と「情報セキュリティ」との間, および「情報セキュリティ」と「ICT 利用のモラルとリスク」との間に有意な差があり (有意水準 1%), 「ICT 利用のモラルとリスク」がもっとも重要であると考えたことがわかる。

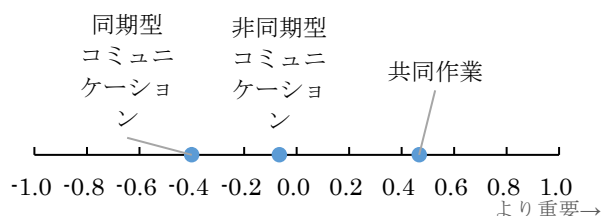


図 3 オンラインコラボレーション力の子ノードの重要度の比較結果

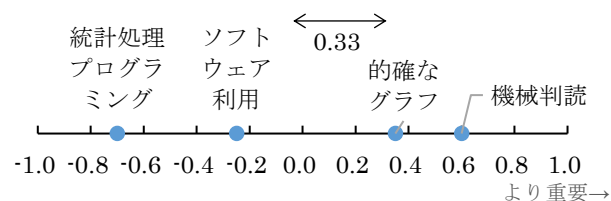


図 4 データ利活用力の子ノードの重要度の比較結果

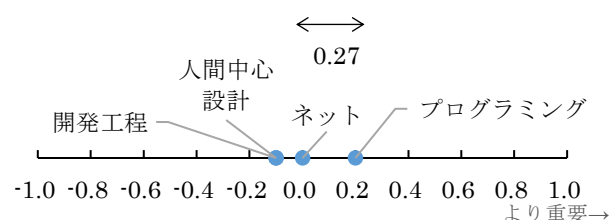


図 5 情報システム開発力の子ノードの重要度の比較結果

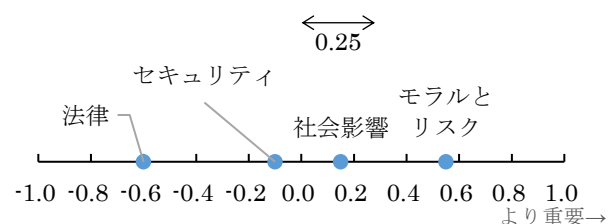


図 6 情報倫理力の子ノードの重要度の比較結果

3.3 考察

一対比較結果の全般として, カテゴリ間および項目間に有意差が見られたことから, 学生による自己評価において, どの項目も同等の重要度で扱うのではなく, 項目ごとに重みをつけて評価結果へ反映させるべきであると考えられる。そのひとつの方法として, 今後, 一対比較に基づいた Analytic Hierarchy Process (AHP)

の利用を計画している。この際には、今回の一対比較における5名よりも多い回答者のデータに基づいた分析を予定する。

図2の4つのカテゴリ間の比較結果において「データ利活用」がもっとも重要と回答されたことから、理系／文系や学部や学科にかかわらず、コンピュータを使ったデータ分析や機械学習およびグラフによる可視化などは重要であると考え、学生がこのカテゴリの項目を重点的に学修していくことが望ましいと考える。

図3とその結果から、オンラインコラボレーション力の下位深さ1の項目間の重要度には差がなく、どれも同程度の重要度として評価する必要があると考える。

4. おわりに

本稿では、大学生のICTに関する知識とスキルである「情報活用学士力」を評価するためのメトリクスを著者らのブレインストーミングによって木構造として体系的に導出し、上位のカテゴリ同士およびカテゴリの子同士を一対比較した。導出した木の構造および枝や葉の項目から、多岐に渡って広く網羅できたと考える。また、各項目から、文系／理系などの分野にかかわらず、オフィススイートやOSの操作以外の、統計処理やプログラミングや人工知能などに関する知識とスキルが必要とされていくことを示せたと考える。すなわち、情報活用学士力として、高校の情報Ⅰまたは情報Ⅱの基礎的知識と基礎的スキルに基づいた、より応用的な知識とスキルが求められていると考える。

多岐にわたっていることは、広く網羅できる一方で、その回答に時間がかかってしまうという欠点がある。この欠点に対しては、木構造で構成した利点をいかし、すべての学生が常に全部の細目を使用するのではなく、各学生が必要に応じて、木構造の中から自身に必要と考える枝や葉を選択して評価していくことで対応できると考える。今後、実際に学生に利用してもらい、このメトリクスの汎用性・効率性・継続性を評価していく。あわせて、一対比較の結果から項目間の重要度に差があることがわかったため、評価の際には重みをつけて評価結果を算出する必要があると考える。このために、重み付けのための質問紙調査を実施し、その結果に基づいてAHPを実施する計画である。

今後は、AHPの実施およびこの木構造の項目のさらなる精査を繰り返し、情報活用学士力の保証に相応しいメトリクスへと改善していく。特に情報分野においては、生成系人工知能などの新たなシステムやサービスが次々と開発／提供されているので、そのような事項への知識やスキルにも柔軟に対応できるメトリクスへと更新していくことを考えている。

謝辞

本研究はJSPS 科研費 JP22K02831 の助成を受けたものである。

参考文献

- (1) 金子大輔, 石田雪也, 小俣昌樹, 吉川雅修, 古賀崇朗: “大学の初年次学生を対象とした情報に関する基礎知識調査の開発と調査結果の分析”, 日本教育工学会論文誌, 40 巻, Suppl. 号, pp.201-204(2017)
- (2) 石田雪也, 金子大輔, 小俣昌樹, 吉川雅修, 古賀崇朗, “初年次情報科目と高校情報Ⅰの学習指導要領の比較,” リメディアル教育学会第 17 回全国大会, pp.56-57 (2022)
- (3) 文部科学省, 高等学校学習指導要領 情報科関係資料, https://www.mext.go.jp/a_menu/shotou/zyouhou/detail/mext_01831.html (2023 年 6 月 6 日確認)
- (4) 文部科学省, 高等学校情報科「情報Ⅰ」教員研修用教材 (本編), https://www.mext.go.jp/a_menu/shotou/zyouhou/detail/1416756.htm (2023 年 6 月 6 日確認)
- (5) 文部科学省, 高等学校情報科「情報Ⅱ」教員研修用教材 (本編), https://www.mext.go.jp/a_menu/shotou/zyouhou/detail/mext_00742.html (2023 年 6 月 6 日確認)
- (6) 総務省統計局, 統計表における機械判読可能なデータの表記方法の統一ルールの策定, https://www.soumu.go.jp/menu_news/s-news/01toukatsu01_02000186.html (2023 年 6 月 6 日確認)

人間的力養成科目の体系化に向けた共通性と個別性の抽出

村松慶一^{*1}, 松居辰則^{*2}

^{*1} 早稲田大学グローバルエデュケーションセンター

^{*2} 早稲田大学人間科学学術院

Extraction of Common and Individual Characteristics toward Systematization of Courses for Self-competence Development

Keiichi Muramatsu ^{*1}, Tatsunori Matsui^{*2}

^{*1} Global Education Center, Waseda University

^{*2} Faculty of Human Sciences, Waseda University

The Global Education Center of Waseda University offers educational programs for "self-development" as one of the academic skills that all students should acquire in common. In this educational area, students are required to select courses that are in accordance with their interests and form a cycle of theory and practice. This report describes an attempt to clarify the common and individual characteristics of the courses by expressing the educational goals of the courses as constructs. As a result of analyzing the constructs, they were classified into 16 classes, and the relationship between subjects and classes could be used to characterize the courses.

キーワード: 人間的力養成, 導入教育, 科目の体系化, コンストラクト, 共通性と個別性

1. はじめに

早稲田大学グローバルエデュケーションセンター(以下, GEC)では, 全ての学生が共通に身に着けるべき学術的スキルのひとつとして, 人間的力養成(self-competence development)の教育プログラムを提供している。人間的力養成という概念は, 早稲田大学の教育・研究ポリシーを表現する上で重要であり, 答えのない問題に挑戦する「たくましい知性」を鍛え, 多様な価値観を持つ人々に敬意をもって接し理解する「しなやかな感性」を育むという教育方針として表現されている。また, GECにおける「たくましい知性」と「しなやかな感性」を育成する取り組みは, 正課の科目を起点として正課外の活動まで含めた理論と実践の往還を目指すものである。

これまでに, 人間的力養成のために導入科目を試

行した⁽¹⁾。具体的には, 世間一般, 仕事, 自己に対する態度について学生自身の認識あるいは再認識を促すことをコンセプトとして授業を設計し実施した。これによって, 自身にとって修得の必要な正課の科目と正課外の活動を学生が自ら選択する意識づけができたと考えられる。目指すところである理論と実践の往還の実現には, 正課の科目と正課外の活動を円滑に接続する必要がある。そのため, 人間的力養成科目の共通性や個別性について整理することで, 人間的力養成に関わる正課の科目と正課外の活動を体系化することが課題である。

本報告では, 正課の科目であるキャリア形成, ダイバーシティ, ボランティア, 地域連携, リーダーシップ, ビジネス創出の科目群に焦点を絞り, それらの体系化に向けて実施した調査と分析について述べる。具体的には, 学生のコンストラクトに着目して, 人間的力

量育成科目の担当教員が当該科目の学習活動において学生に関連付けを期待する視点・観点を調査し、その分類を行うことで各科目の共通性と個別性を明確にする。コンストラクトの考え方については、次節で述べる。

2. コンストラクトの考え方

先述の導入科目では、学生自身が世界に対する態度の認識あるいは再認識を行うことで、(1) 学生自身が伸ばしたい人間的力量を自ら認識し、(2) 修得の必要な正課の科目と正課外の活動を学生が自ら選択できるようになることを期待している。このうち(2)が円滑に行われるためには、人間的力量育成科目それぞれに設定している教育目標が学生自身にとってどのような意味を持つのか、言い換えればその科目を受講することで学生自身にどのような成長が期待できるのかということが明確である必要がある。また、受講による自身の成長を容易にイメージすることができれば、翻って(1)の促進も期待できる。教育目標はしばしば行動目標として表現されるが、導入科目では世界に対する態度、言い換えれば世界の捉え方に注意を向けさせていることから、人間的力量育成科目の教育目標も同様な表現がなされれば接続性が高くなると考えられる。

世界の捉え方を表す考え方の一つにコンストラクトがある。Georg A. Kelly によるパーソナル・コンストラクト理論^②では、コンストラクトとは世界を構成する概念を指す。この理論は、人間はさまざまな事象を経験し、事象間の類似性や差異を認知し、現象についての構成概念（コンストラクト；construct）を形成し、それに基づいて事象を予測しようとする、という仮説に基づいている^③。構成概念は双極的（二分法的）概念であり、それらが人間の解釈システムを構成すると理解される。この理論を背景とした面接手法を発展させた評価グリッド法^{④、⑤}は、人間の解釈システムのうち評価に関わる評価構造を階層的に可視化する手法として建築分野や心理学分野で多く活用されている。

人間的力量育成科目の教育目標を表現する際にも、コンストラクトの考え方を援用することできると思われる。すなわち、各科目の学習内容に関連するコン

ストラクトの組み合わせで、行動目標とは別の形で教育目標を表現するということである。行動目標は多くの場合で行動とその対象が合わせて表現されるため、対象領域の文脈に強く依存している。一方で、コンストラクトは双極的あるいは二分法的な形容語の対という形をとり、行動目標に比べて抽象度の高い表現となる。また、コンストラクトは世界の認識を構成する概念であるという点で、導入科目で認識あるいは再認識した世界に対する態度、言い換えれば世界の捉え方との類似性が高いと考えられる。それゆえ、導入科目と人間的力量育成科目の接続性を高めることが期待される。

3. 方法

3.1 コンストラクトの収集

導入科目である「人間的力量ファンダメンタル」のレビューシートからそれぞれの人間的力量育成科目に関連するコンストラクトの収集を行う。レビューの対象となる学習活動は、1 つの論題に対して肯定、否定の両方の立場から立論、反論、主張を作成するセルフ・ディベートである。2021 年度に実施した導入科目の試行では対面形式の授業の中でディベート形式のディスカッションを行っていた^⑥が、2022 年度から授業がオンデマンド形式に変更されたことに伴ってディベート形式のディスカッションはセルフ・ディベートに置き換えられた。

それぞれの科目群とセルフ・ディベートの論題は表 1 の通りである。これらの論題は「正解のない問題」を念頭に各科目の担当教員との協議によって作成され、一般的な意見としては肯定とも否定ともつかないような論題とすることで、特定の文脈に依存した主張が出されやすくなることを意図したものである^⑦。受講者は各論題について肯定、否定それぞれの立場から立論、反論、主張を作成し、授業支援システム上のフォーラムに投稿することが求められる。また、他の受講者の投稿に対して返信機能でコメントすることで、それぞれの気づきを共有することも求められる。レビューシートでは、この学習活動を振り返り論題ごとに気づいた自身の「思い込み」を理由とともに記述することが求められる。コンストラクトは相対する形容語を両極

とした次元として捉えられるため、レビューシートに記述された気づきの内容を学生の視点や観点の変化をその次元内の変化と見做すことで、コンストラクトを解釈できると考えられる。あるいは、次元そのものに気づく場合であっても、気づきの内容から学生の視点や観点の変化に着目することでコンストラクトを解釈可能であると考えられる。これらの解釈を担当教員が行い、レビューシートから形容語対として表現されるコンストラクトを収集する。

表 1 科目群に関連するセルフ・ディベートの論題

科目群	論題
キャリア形成	具体的なキャリア目標を持つべきか、否か
ダイバーシティ	「みんなちがって、みんないい」という価値観は、競争社会において受容可能か、否か
ボランティア	ボランティアは自己満足か、否か
地域連携	地域連携による地域活性は持続可能か、否か
リーダーシップ	リーダーシップをとるのは牽引力のある人(たち)だけであるべきか、否か
ビジネス創出	ビジネス創出は大学や企業等のシーズに基づくべきか、否か

3.2 共通性と個別性の抽出

ひとりディベートの学習活動に対するレビューシートから収集されたコンストラクトについて、人間的力量育成科目との関連性について調査を行う。具体的には、人間的力量育成科目の担当教員に依頼し、担当科目の教育目標や学習内容に基づいて受講生に期待する視点・観点をコンストラクトの一覧から選択してもらう。その際に、受講生があらかじめ当該の視点・観点を備えていて、その次元内での立ち位置を変えながら学習活動に取り組んで欲しいといった場合と、学習活動を通して当該の視点・観点を新しく発見して欲しいといった場合に分けて回答を得る。ただし、本報告

ではそれらを区別した議論は今後の課題とし、以降の分析でそれらの区別を行わず単に科目と関連する視点・観点として扱うこととする。また、コンストラクトの一覧に記載されていない視点・観点があれば形容語対の形で新しく加えてもらうが、本報告では分析の対象外とする。

得られた回答データについて、科目と関連しないものを「0」、関連するものを「1」としてダミー変数を割り当てて潜在クラス分析を行うことでコンストラクトの分類を行うとともに、科目ごとの正反応確率（ダミー変数「1」の反応確率）からそれぞれの科目に特徴的なコンストラクト群を明らかにすることで科目の共通性と個別性と検討する。

4. 結果

4.1 コンストラクトの収集

2022 年度の春クォータから冬クォータのそれぞれで開講された「人間的力量ファンダメンタル」のセルフ・ディベートに対するレビューシートを分析の対象とした。セルフ・ディベートは論題をふたつずつ3回に分けて実施され、のべ455件のレビューシートが得られた。記述された学生の気づきの内容について、学生の視点や観点の変化を形容語対で表現される次元として担当教員が解釈を行った。その結果、112の形容語対がコンストラクトとして収集された(表2)。文脈によって異なる解釈が可能な形容語対については、括弧書きで意味が追記されている。

レビューシートにおいて、コンストラクトを表すような形容語が用いられずに学生の視点や観点の変化が記述されている場合には、担当教員の解釈によってコンストラクトを推測した。例えば、「短期的」や「長期的」といった形容語が用いられていないものの「時間軸を考慮していないことに気づいた」というように時間的な視点・観点について言及しているものは、担当教員の解釈によって「短期的-長期的」のコンストラクトに該当すると推測された。また、コンストラクトを表す形容語対のうち、一方の極についてのみレビューシートに記述されている場合には、担当教員の解釈によってもう一方の極にあたる形容語を推測した。例え

ば「専門的」という形容語のみが用いられているものは、もう一方の極として「汎用的」や「一般的」の形容語が推測された。そのため、収集されたコンストラクトの中には類似した意味の形容語対が含まれている。

表 2 収集されたコンストラクトの一覧

具体的-抽象的	過程的-結果的	狭い-広い（人間関係）
絶対的-相対的	持続的-断続的	狭い-広い（視野）
経験的-思索的	目的的-手段的	小さい-大きい（枠組み）
専門的-一般的	動機的-結果的	微視的-巨視的
専門的-汎用的	可能な-不可能な	小さい-大きい（可能性）
実践的-理論的	安心な-不安な	小さい-大きい（規模）
現実的-理想的	消費者の-生産者の	効率的-非効率的
善い-悪い	得意な-不得意な	低コストな-高コストな
良い-悪い	受給者の-提供者の	生産的-非生産的
正当な-不当な	成功の-失敗の	效果的-逆効果
理性的-感情的	活発な-緩慢な	低収益な-高収益な
ポジティブな-ネガティブな	個性的-没個性的	偶発的-計画的
高い-低い（価値）	前向きな-後ろ向きな	可変的-不変的
トップダウン的-ボトムアップ的	成長的-衰退的	状況依存的-普遍的
利己的-利他的	個性的-類型的	柔軟的-硬直的
対立的-協調的	努力的-怠惰的	流動的-固定的
競争的-協力的	縦の-横の（関係）	不安定な-安定な
個別的-普遍的	満足な-不満な	変動的-安定的
若年の-老年の	能動的-受動的	新規の-既存の
若者の-年配の	主観的-客観的	先進的-後進的
一方向の-双方向の	自由な-不自由な	魅力的-退屈な
強い-弱い	自由な-統制的	新しい-古い
裕福な-貧困な	選択的-強制的	革新的-保守的
独裁的-民主的	自発的-強制的	挑戦的-堅実的
独断的-合議的	生得的-経験的	おもしろい-つまらない
極端な-中庸な	主体的-従属的	有限な-無限な
ローカルな-グローバルな	積極的-消極的	個人的-社会的
地方の-中央の	先天的-後天的	本質的-表面的
地域間の-地域内の	素質的-能力的	個人的-組織的
多数派の-少数派の	才能的-経験的	個別的-集团的
共通の-独自の	内面的-外面的	部分的-全体的
特殊的-普遍的	一様な-多様な	構造的-要素的
標準的-例外的	単一的-複合的	私的-公的
普通な-特別な	一面的-多面的	内的-外的
直接的-間接的	単純な-複雑な	定量的-定性的
顕在的-潜在的	一元的-多元的	有形の-無形の
閉鎖的-開放的	択一的-重疊的	
短期的-長期的	狭い-広い（選択肢）	

4.2 共通性と個別性の抽出

本報告ではボランティア、ビジネス創出、地域連携の科目群に限定して、収集されたコンストラクトと人間的力量育成科目との関連性についての調査を行った。その結果、5名の教員が担当する18科目についての調査データが収集された。潜在クラス分析を行う前処理として、どの科目からも選択されなかったコンストラクトや複数の科目で全く同じように選択されたために区別のつかないコンストラクトを除外するために階層的クラスター分析(ウォード法, ユークリッド距離)を行った, 分析には HAD18.0⁽⁶⁾を用いた。距離が0でク

ラスターを形成した39のコンストラクトを抜き出し、そのうちの科目からも選択されなかった14のコンストラクトは除外した。残りの25のコンストラクトは、複数の科目で全く同じように選択されたために区別のつかないものであったため、クラスターごとに代表となるコンストラクトを残して除外した。

前処理で得られた80のコンストラクトについて潜在クラス分析を行った。分析には HAD18.0⁽⁶⁾を用いた。潜在クラス分析では、通常は適合度指標を比較してクラス数を決定する。しかし、本報告ではコンストラクトの分類による各教科の特徴付けに主眼をおいているため、カテゴリごとの正反応確率が1となるクラスが、各カテゴリに1つ以上となるような最小のクラス数を探索した。すなわち、クラス内のすべてのコンストラクトが1つの教科と関連するクラスを、各教科が少なくとも1つもつようにモデルを決定した。その結果、16クラスのモデルを採用した。モデルの統計学的な妥当性は度外視しているため、適合度指標の比較については割愛する。

各クラスに所属するコンストラクトの数と代表例は表3の通りである。クラス1からクラス16に所属するコンストラクトの数は順に2, 1, 9, 4, 6, 8, 3, 4, 10, 7, 3, 4, 7, 4, 3, 5であった。また、所属するコンストラクトの代表例は順に次の通りであった。クラス1: 安心な-不安な, 私的-公的, クラス2: 個人的-組織的, クラス3: 個性的-没個性的, 素質的-能力的, など, クラス4: 新規の-既存の, 先進的-後進的, など, クラス5: 経験的-思索的, 挑戦的-堅実的, など, クラス6: 善い-悪い, 多数派の-少数派の, など, クラス7: 生産的-非生産的, 效果的-逆効果, など, クラス8: 効率的-非効率的, 低収益な-高収益な, など, クラス9: 正当な-不当な, 持続的-断続的, など, クラス10: 具体的-抽象的, 実践的-理論的, など, クラス11: 動機的-結果的, 構造的-要素的, など, クラス12: 若者の-年配の, 本質的-表面的, など, クラス13: 自由な-不自由な, 自発的-強制的, など, クラス14: 共通の-独自の, 個人的-社会的, など, クラス15: 絶対的-相対的, 狭い-広い（選択肢）, など, クラス16: 対立的-協調的, 個別的-普遍的, など。

所属するコンストラクト群に基づいてクラス名は順に次のように命名された。クラス 1：社会的関係の視点、クラス 2：個人と組織の視点、クラス 3：個性の視点、クラス 4：前向きと後ろ向きの視点、クラス 5：発想と実践の視点、クラス 6：意見共有の視点、クラス 7：生産性の視点、クラス 8：効率性の視点、クラス 9：活動の意味の視点、クラス 10：課題把握の視点、クラス 11：因果関係の視点、クラス 12：価値観受容の視点、クラス 13：取り組み姿勢の視点、クラス 14：個人と社会の視点、クラス 15：判断基準の視点、クラス 16：相互理解の視点。科目ごとの正反応確率を表 3 に示す。紙面の都合上クラス 1 から 8 を上段、クラス 9 から 16 を下段に分割している。

5. 考察

潜在クラス分析によって、導入科目のレビューシートから得られたコンストラクトは 16 クラスに分類さ

れた。各科目が 1 つ以上のクラスによって特徴付けられるようにモデルを選択したため、科目の異同はどのクラスと関連しているかという点から説明することができる。すなわち、同じクラスと関連することを共通性、他とは異なるクラスと関連することを個別性と思ふことができる。表 3 の正反応確率が科目とクラスの関連性の高さを示す。ここで正反応確率が 1 であるもののみを抽出すると、例えば「体験の言語化」という科目はクラス 2：個人と組織の視点、クラス 14：個人と社会の視点という 2 つのクラスで特徴付けられる。同様に「起業特論 B」という科目はクラス 2：個人と組織の視点、クラス 13：取り組み姿勢の視点、クラス 15 判断基準の視点という 3 つのクラスで特徴付けられ、「地域連携」という科目はクラス 1：社会的関係の視点、クラス 2：個人と組織の視点、クラス 7：生産性の視点、クラス 10：課題把握の視点、クラス 12：価値観受容の視点、クラス 14：個人と社会の視点

表 3 科目ごとの正反応確率

	C1: 社会的関係	C2: 個人と組織	C3: 個性	C4: 前向きと後ろ向き	C5: 発想と実践	C6: 意見共有	C7: 生産性	C8: 効率性
ボランティア論	0	0	.000	.500	0	.735	1.000	.250
グローバル社会貢献論	0	1.000	.211	.500	.322	.873	1.000	.750
パラスポーツとボランティア	0	0	.774	.250	0	.490	0	.250
グローバルヘルス	0	1.000	.211	.500	.000	.735	0	.000
体験の言語化	.500	1.000	.000	0	.337	.372	0	0
ボランティアから見る世界	0	0	0	.250	.337	.739	.333	0
狩猟と獣害対策論	.500	0	0	0	.486	1.000	0	.750
狩猟と地域おこしボランティア	.500	0	.000	.250	1.000	1.000	.333	.750
イノベーション人材になるためのコーチング研修	.500	0	.338	0	.000	.490	0	.250
起業特論B	0	1.000	.000	.750	.493	.627	0	.000
Advanced Course on Entrepreneurship D	0	0	.225	1.000	.514	.000	0	.250
地域連携学	1.000	1.000	0	.500	0	.000	1.000	.500
地域連携演習：基礎演習	0	0	0	.000	0	0	.333	.000
地域連携演習：実践演習	0	0	0	1.000	0	0	0	1.000
地域連携演習：ケーススタディ	0	0	0	1.000	0	0	.667	.750
地域連携演習：コミュニティ・インタラクション	0	1.000	.113	1.000	0	0	0	1.000
地域連携演習：レポート演習	0	0	0	0	0	0	0	0
地域連携演習：プレゼンテーション演習	0	0	0	0	0	0	0	0

	C9: 活動の意味	C10: 課題把握	C11: 因果関係	C12: 価値観受容	C13: 取り組み姿勢	C14: 個人と社会	C15: 判断基準	C16: 相互理解
ボランティア論	1.000	.429	1.000	0	.339	.750	0	.800
グローバル社会貢献論	1.000	.571	.000	0	.711	.750	.333	.829
パラスポーツとボランティア	.678	.429	0	.250	.158	.750	.333	1.000
グローバルヘルス	.379	.286	0	.250	.000	.750	0	.800
体験の言語化	.398	.143	.354	.250	0	1.000	.333	.415
ボランティアから見る世界	.000	1.000	.354	0	.395	.500	1.000	.509
狩猟と獣害対策論	.000	1.000	.354	0	0	.250	1.000	.000
狩猟と地域おこしボランティア	.226	1.000	0	.500	.000	.250	.667	.000
イノベーション人材になるためのコーチング研修	.407	.286	0	0	.794	.250	1.000	.579
起業特論B	.385	.143	.293	0	1.000	.750	1.000	.378
Advanced Course on Entrepreneurship D	.000	.571	0	.250	.400	.000	1.000	.672
地域連携学	0	1.000	0	1.000	0	1.000	0	0
地域連携演習：基礎演習	0	1.000	0	1.000	0	.750	0	0
地域連携演習：実践演習	0	1.000	0	1.000	0	1.000	0	0
地域連携演習：ケーススタディ	0	1.000	0	1.000	0	1.000	0	0
地域連携演習：コミュニティ・インタラクション	0	1.000	0	1.000	0	1.000	0	0
地域連携演習：レポート演習	0	1.000	0	.750	0	.250	0	0
地域連携演習：プレゼンテーション演習	0	1.000	0	.750	0	.500	0	0

という 6 つのクラスで特徴付けられる。これらの科目とクラスの関連から、3 つの科目の共通性はクラス 2：個人と組織の視点であるといえる。さらに「体験の言語化」と「起業特論 B」の比較から、「体験の言語化」の個別性はクラス 14：個人と社会の視点であり、「起業特論 B」の個別性はクラス 13：取り組み姿勢の視点、クラス 15 判断基準の視点であるといえる。本報告で調査した科目は「ボランティア」「ビジネス創出」「地域連携」の科目群に属している。それぞれの科目群内では関連するクラスが共通し、逆に科目群間では関連するクラスが異なる傾向がみられることから、潜在クラス分析によるコンストラクトのクラス分類と科目の特徴付けは一定の妥当性を有していると考えられる。

各教科と関連するコンストラクトのクラスは、担当教員が受講生に望む変化、すなわち科目として意図している受講生の成長を表しているため、教育目標のひとつの表現と見做せる。したがって、この調査と分析によって教育目標の共通性と個別性を明確にすることができたと考えられる。教育目標を行動目標として表現する場合には、対象領域に依存した行動について言及することになる。一方で、視点・観点として教育目標を表現した場合には、対象領域には依存しにくく、科目の共通性と個別性を明確化することができる。その意味で、本報告の調査と分析は新しい分析手法として有用であると考えられる。

この手法の成否でもっとも重要な点は、コンストラクトが適切にクラスに分類されることである。それについては、本報告で分類した 16 クラスでは十分にコンストラクトを分類しきれていない可能性があり、今後の検討課題である。コンストラクトと科目の関係を取得する調査において、サンプル数を増やすことでより適切なコンストラクトのクラス分類が得られることが期待される。

6. おわりに

本報告では、人間的力養成科目とコンストラクトの関連性を調査し、その分類を行うことで各科目の共通性と個別性を明確にした。調査のサンプル数を増やすことで、人間的力養成科目全体と共通性と個別性

を明確化することが今後の課題である。さらに、正課外の活動にも調査の範囲を広げることで、正課の科目と正課外の活動を円滑に接続するための、科目間の関係性を図示するようなツールの作成を目指す。それによって、修得の必要な正課の科目と正課外の活動を学生が自ら選択することが容易になると考えられる。

謝辞

早稲田大学グローバルエデュケーションセンター人間的力養成教育部門の先生方には、調査の設計と実施について有益な助言をいただきました。特に、岩井雪乃先生、加藤基樹先生、島岡未来子先生、兵藤智佳先生（五十音順）には調査協力および助言をいただきました。記して謝意を表します。

参 考 文 献

- (1) 村松慶一，松居辰則：“人間的力養成のための導入教育の設計と試行”，教育システム情報学会研究報告，Vol.37, No. 1, pp.44-48 (2022)
- (2) G. A. ケリー(著)，辻平治郎(訳)：“パーソナル・コンストラクトの心理学 第1巻 理論とパーソナリティ，第2巻 臨床診断と心理療法”，北大路書房，京都 (2016)；Kelly G. A.: “The Psychology of Personal Constructs, Vol.1. A Theory of Personality, Vol. 2. Clinical Diagnosis and Psychotherapy”, W. W. Norton, New York, (1955)
- (3) 若林明雄：“George A. Kelly の個人的構成概念の心理学”，心理学評論，Vol.35, No.3, pp.311-338 (1992)
- (4) 讃井純一郎，乾正雄：“レパートリーグリッド発展手法による住環境評価構造の抽出-認知心理学に基づく住環境評価による研究 (1) -”，建築学会計画系論文集，No.367, pp.15-22 (1986)
- (5) 宇治川正人，丸山玄，讃井純一郎：“電子メールを用いた評価グリッド法の開発魅力あるガソリンスタンドの条件に関する調査研究”，建築学会計画系論文集，No.518, pp.75-80 (1999)
- (6) 清水裕士：“フリーの統計分析ソフト HAD：機能の紹介と統計学習・教育，研究実践における利用方法の提案”メディア・情報・コミュニケーション研究，Vol.1, pp.59-73 (2016)

L-VEIGe : 誤答画像生成を活用した語彙学習支援システム

— 画像の示唆性に対する指標の検討 —

杉田 一樹^{*1}, 太田 光一^{*1}, 谷 文^{*1}, 長谷川 忍^{*1}

^{*1} 北陸先端科学技術大学院大学

L-VEIGe: Vocabulary Learning Support System

Using Error Image Generation

-A study of criteria for image suggestibility -

Kazuki SUGITA^{*1}, Koichi OTA^{*1}, Wen GU^{*1}, Shinobu HASEGAWA^{*1}

^{*1} Japan Advanced Institute of Science and Technology

The L-VEIGe (Learning-Vocabulary Error Image Generation) system for supporting vocabulary learning has been shown to promote learners' introspection and prevent repeated mistakes by visualizing errors through image generation. Learners' impressions and memorability are considered important criteria for the generated images. This paper examines a quantitative method for learning efficiency and addresses the problem of feature disappearance in the generated images that is necessary to understand the vocabulary in the errors.

キーワード: 語彙学習支援システム, 第二言語習得, 誤答画像の示唆性, 生成画像評価指標

1. はじめに

英語語彙学習は英語を第二言語とする L2 学習者に於いて重要とされ, 語彙はコンテキストから学習する方法が有効であるとされている⁽¹⁾. 語彙学習では, 画像を用いた研究が多く存在し, 画像から対応する文を自動で生成するイメージキャプショニングを活用した語彙学習支援システムが存在する⁽²⁾. ここでは, 画像をコンテキストと見做すことが可能だとし, 膨大な読書量に対する画像の代替の可能性を示している. しかしながら, これらは学習時の誤答については言及されていない. 一方, 英語学習に於いて, 誤った認識のまま繰り返し用いる事によることに依る化石化「fossilization」の問題がある⁽³⁾. 化石化では, その誤りが一度定着すると, その修正は困難であるとされている. つまり, 学習者の語彙に対する誤りを効果的に認識させることが重要であると捉えられる. そこで筆者らは, 英語語彙補充問題を基本とし, 学習者の誤答

を画像生成により可視化することにより学習者の誤答に対する内省を促す, 語彙学習支援システム L-VEIGe(Learning-Vocabulary Error Image Generation)を提案している.

本稿では, 提案手法と評価実験の概要について述べ, 生成される画像の示唆性に於ける想起性に関する課題である特徴消失について議論し, その解決策として想起性の定量的評価指標を提案する⁽⁴⁾.

2. 提案手法

ここでは, L-VEIGe の構成について述べる. 既存のデータセットである MS-COCO データセット⁽⁵⁾を活用する事により, 以下のタスクを定義し, ①～③を実現した.

- ① 質問生成タスク: 学習用教材として網羅的である問題の自動生成を実現,
- ② 回答支援タスク: 語彙学習に於いて近似したスベルが間違いを誘発するとされることから, 近

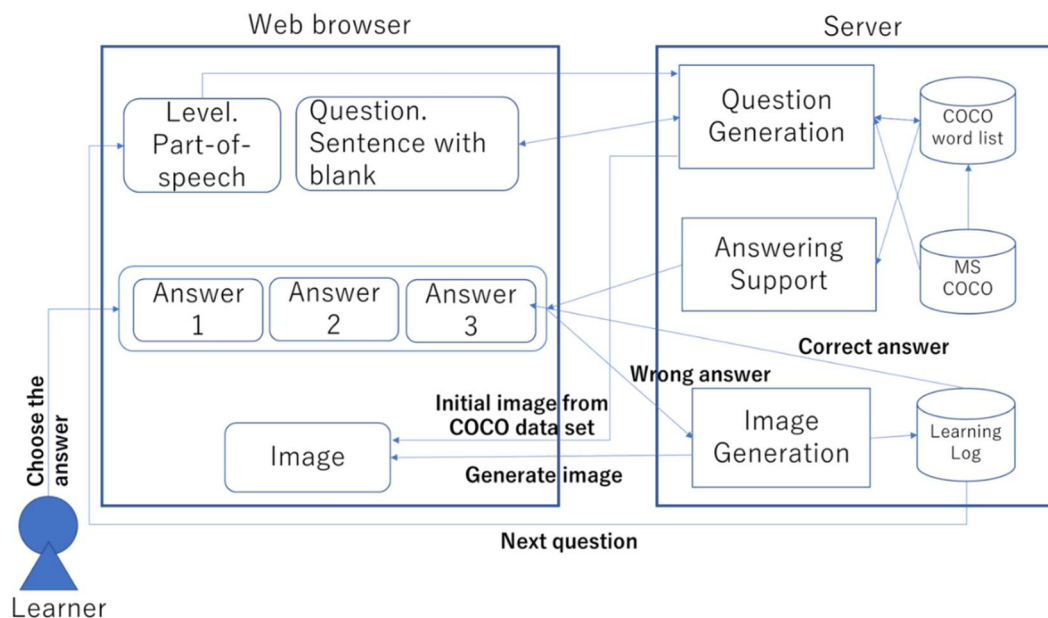


図 1 L-VEIGe ユースケース図

似したスペルを有する語彙を学習対象語彙の回答の候補として自動的に抽出し、誤答を促す環境を構築。

- ③ 画像生成タスク：文に応じた画像を生成可能とする DALL-E⁽⁶⁾を活用し、学習者の誤答に応じた画像生成を実現。

これら①～③は、図 1 に対応する。図 2 は学習者の誤答に応じ、画像が変化した様子を表す。ここでは、学習者は、初めに左画像とその画像に対応する文「Two sheep and a () stand next to a fence in the yard」に対し、正答が ram となるが、学習者が tram を選択した為、誤答を含む文に対応した画像が提示される。

3. 評価実験

繰り返しの誤答の回避、誤答した語彙に対する理解を評価する為に、英語を第二言語とする TOEIC 500～900 点を有する、8 カ国 24 名の理系大学院生を対象とし実験を行った⁽⁷⁾。実験では、システムエラーに依るデータ欠損の為、被験者 22 名から得られたデータを評価対象とした。被験者は、誤答時の画像生成を有する Condition1 と画像生成を有さない Condition2 の両方に依る学習を連続で行い、2 つの条件での連続した学習を 1 タームとし、1 ターム目と同じ問題、条件を繰り返す合計 2 タームに依る実験を行った。被験者は同一の問題が与えられ、条件の順序、問題の順序に

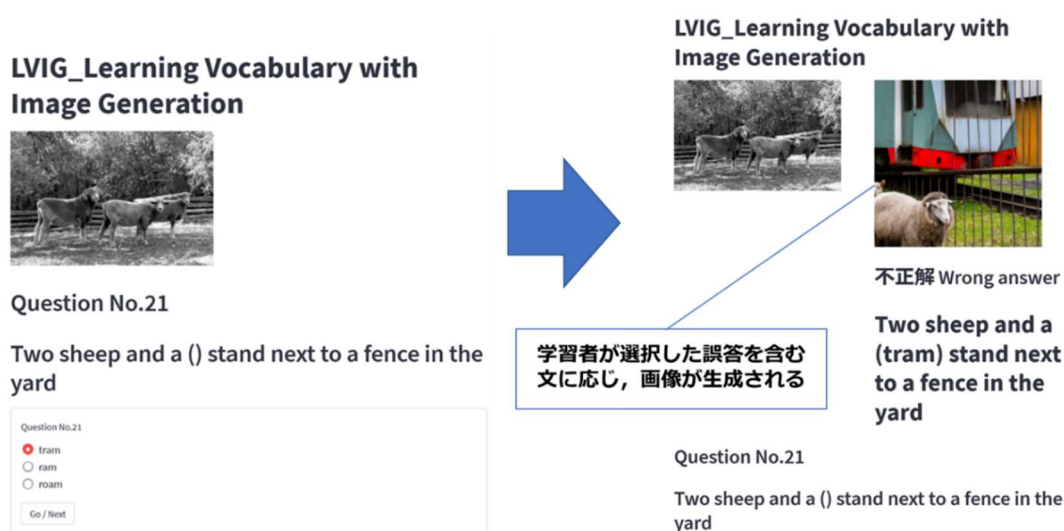


図 2 L-VEIGe システム操作例

依る効果を解消した被験者内実験に依るカウンターバランスを取った。又、学習対象者が繰り返しの誤答を避けられたか判定する為、意図的に 20% すべての選択肢が誤りである問題を混在させた。又、生成画像を評価する上で、画像に対する、印象、想起性の二つの評価軸を定義し、画像の示唆性と定義した。被験者は初期画面に表示される画像、誤答時に生成される画像に対して、

- A. 初期画像の想起性：初めに表示される画像は文を想起させるものか、
- B. 生成画像の想起性：生成された画像は文を想起させるものか、
- C. 生成画像の印象：生成された画像は印象に残るか、

の 3 点について 5 段階評価を行った。又 3 日後に被験者が誤答した語彙に対し理解度を問うアンケートを行った。

3.1 繰り返しの誤答の回避に対する評価

ここでは、学習者が繰り返しの誤答を避けられたか否かの評価を行い、ターム 1、ターム 2 に於ける、Condition1, 2 による誤答の繰り返し数を基に評価を行う。ここでは、評価対象者 22 人の結果に対して、ノンパラメトリックな手法である Wilcoxon の符号順位検定を行う。ここでは、帰無仮説 H_0 を Condition1 と Condition2 の繰り返しの誤答の回数に差がないとする。検定により導出された 2 群間のデータの差の絶対値を W と定める。サインランク表を用いた検定により、片側有意水準 $\alpha=0.05$ 、棄却限界値 53、 $W=45$ 、 $W < 53$ より、帰無仮説を棄却する。図 3 に示す度数分布表より、提案手法である画像生成を伴うモデルが繰り返しの誤答を防ぐうえで有効である事を示した。

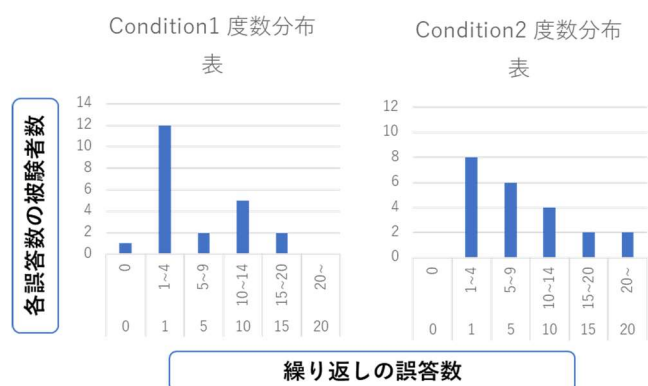


図 3 印象と繰り返しの誤答数との相関関係

3.2 生成画像の特性と繰り返しの回避の関係

学習者に依る定性評価による、A. 初期画像の想起性、B. 生成画像の想起性、C. 生成画像の印象のうち、誤答に関する評価である、B,C について、繰り返しの誤答数との相関をノンパラメトリックな指標である、Spearman の順位相関係数に依る検定を行った。想起性と繰り返しの誤答数についての検定では、帰無仮説 H_0 を想起性と繰り返しの誤答数に相関関係はないとする。図 4 左では、y 軸を画像の想起性、x 軸を誤答の繰り返し回数を表している。図 4 右ではスピアマンの順位相関係数である ρ 、P 値を示す。ここでは、 $\rho = 5.57$ 、 $P = 0.99 > 0.05$ より帰無仮説を採択する。つまり、想起性と繰り返しの誤答数については有意な相関を得られなかった。印象と繰り返しの誤答数についての検定では、帰無仮説 H_0 を印象と繰り返しの誤答数に相関関係はないとする。図 5 では、 $\rho = -0.24$ 、 $P = 0.03 > 0.05$ より帰無仮説を棄却する。ここでは、印象と繰り返しの誤答数に於いて、弱い負の相関が得られた。これらの両者の比較では、生成画像の想起性より、印象の度合いがより繰り返しの誤答に影響を与えていることが伺える。

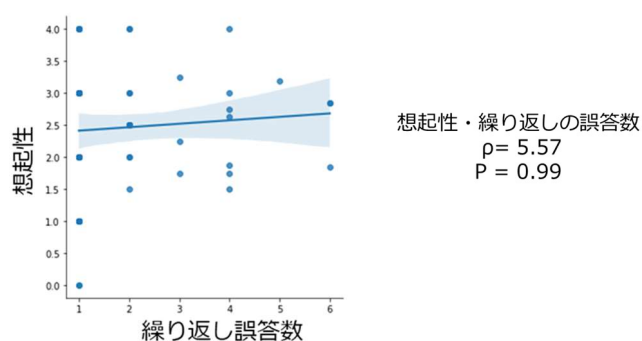


図 4 想起性と繰り返しの誤答数との相関関係

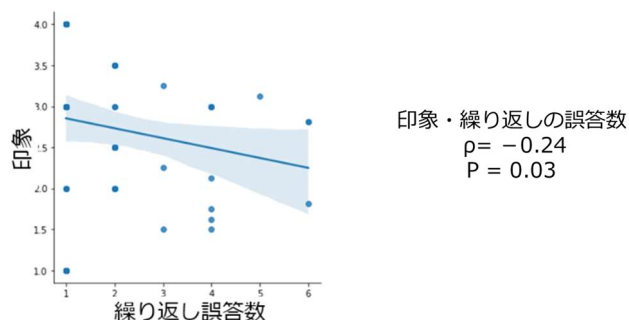


図 5 印象と繰り返しの誤答数との相関関係



図 6 特徴消失した生成画像例

3.3 誤答に対する理解に対する評価

遅延テストでは、誤答に対する理解度として、被験者ごとに誤った語を提示し、それらの語彙に対して、“既に知っている”、“想像できる”、“想像できない”のいずれかを回答する方式で行った。被験者に於ける各コンディションの理解度の評価として、“想像できる”と回答した語彙数から“想像できない”と回答した語彙数の差を誤った語の理解度の評価値とした。先述した繰り返しの誤りの回避に対する評価同様、両者に於ける誤答の理解度に対する評価として、ノンパラメトリックな手法である Wilcoxon の符号順位検定を行う。ここでは、帰無仮説 H_0 を Condition1 と Condition2 の繰り返しの誤答の回数に差がないとする。サインランク表を用いた検定により、片側有意水準 $\alpha=0.05$, 棄却限界値 41, $W=49$, $W>41$ より、帰無仮説を採択する。つまり、誤答した語彙の理解度について、有意な差は見られなかった。繰り返しの誤りの回避に対する評価の結果、今回の結果より誤答への効果的な印象を与え、繰り返しの誤答を防げたが、誤った語を理解するという点に於いては、不十分である可能性を有していることが伺える。

3.4 生成画像と誤答の理解の関係

3.4.1 特徴消失問題

誤った語に対する被験者の回答で、“想像できなかった”という回答が 2 回以上選ばれた語彙について生成画像より分析を行う。“yawn”, “ecstatic”, “fence” を例示する。図 6 では、それら誤った語を入力とした際に生成される画像、及び問題を示している。青色で記された語彙は、その時の正答を示す。図 6 下部の値については、それらの語の誤りの“繰り返し回数”、“想

起性”、“印象”の値を示す。ここでは、これらの生成画像に於いて、画像とセンテンスの対応を表す、“想起性”に於いて 2 以下の低い数値を示した。ここでの“想像できなかった”とされる画像に共通している点として、図 6 より空所として補充した語彙が画像に反映されない特徴消失が発生していることが伺える。

3.4.2 特徴消失問題に依る誤解の誘発

特徴消失によって、明示的に誤答の意味が画像に示されない場合、学習者は、誤答であることを正しく認識できないまま語彙を捉えることが危惧される。図 7 では、学習者が左図の画像に対する回答として、「decorative」を選択することによって、生成された画像を示している。一方、右図では、生成された画像は四角枠で示す通り黒い板が生成され、これによって、学習者は「decorative」の意味を黒い板と関連して記憶する可能性を有する。ここでは、「decorative」の本来の意味は「装飾的な」等の意味を有することから異なる認識を与える原因となりうる。



図 7 誤解を誘発する生成画像例

a:学習者の誤答に応じた生成画像



b:学習者の誤答を含む問題文

Two sheep and a (tram) stand next to a fence in the yard



c:生成されたキャプション

1. A couple of groups of brown sheep standing on the side of a track.
2. A sheep looking at a water trough while walking with a person behind it. and the
3. A flock of white and green sheep walk by a barn and two black sheep looking out
4. A group of brown and black sheep looking toward a group of white sheep outside a barn
5. The two sheep are standing next to others on the side of a field.

d: b - cの近似度

1. 0.72919154
2. 0.6544069
3. 0.7812311
4. 0.72477007
5. 0.891336

図 8 コンテキストの類似度の推定

4. 誤答に対する想起性の指標

評価実験より、誤答時の効果的な印象付けが期待でき繰り返しの誤答を回避する上で有効であることを示す一方で、印象と想起性の二つの軸により定義した画像の示唆性について、特に想起性について特徴消失の問題を有する。誤答への概念理解という観点では、生成される画像が誤答を理解するに足る情報を有している必要があり、従来の手法では、生成される画像の特徴の消失に対する定量的な基準がない為に不十分である。ここでは想起性に対する定量的な評価基準として、コンテキストの類似度、誤り概念の類似度、の二つの評価尺度を定義し、それぞれの方法について提案する。

4.1 関連研究

画像から生成されたキャプションに対する定量的な評価指標として、CLIP スコア⁽⁹⁾が存在する。CLIP スコアでは、画像とキャプションをそれぞれベクトル化し、その一致度を指標とすることにより、画像とキャプションの評価を可能とする。従来の手法では、正解となるキャプションと生成されたキャプションの差異に依る評価が一般的であったが、CLIP スコアでは、正解となるキャプションを必要とせず評価が可能となる。そのため、生成画像と文の全体の一致という観点では評価が可能であるが、ベクトル化された表現は可読性が低いために、特定の語彙の特徴の評価が困難である。

4.2 想起性に対する定量的評価手法

本研究では、文の特徴、誤答となる学習対象語の特徴に注目し、それらの特徴を画像が含意しているか評

価するにあたり、以下の 2 つの評価尺度を定義する。

1. コンテキストの類似度：語彙を学ぶ上でコンテキストが重要と捉えられるため、生成画像に含意する情報に十分なコンテキストを有しているか評価する。ここでは、画像に含まれる情報を自然言語化し、対象となる文と抽出した文の一致度が閾値以上であるとき、生成画像は十分な特徴を有すると定義する。
2. 誤り概念の類似度：学習者が誤答した際に選択した語彙を学習対象語と捉え、その語彙を理解するに足る情報が生成画像に有しているか評価する。誤答となる学習対象語と 1 のプロセスで生成された文中に含まれる単語の一致度を測定し、特定の語彙が閾値以上であるとき、生成画像十分な情報を有すると定義する。

上記 2 点を満たす生成画像を想起性に於ける定量的評価値として定める。

4.2.1 コンテキストの類似度

画像と文の全体の一致度を測定する上で、初めに、画像から自然言語化する方法として、イメージキャプションニングを活用し、画像の特徴を記述した文を得る。ここでは、エンコーダーとデコーダーを事前学習した CLIP⁽⁹⁾をベースに構築された、antarctic-captions⁽¹⁰⁾をモデルとして適用した。図 8 は学習者の誤答に応じた生成文と、生成画像から抽出したキャプションの一致度を表す。a は、対象となる画像を表し、b は、生成された画像の基となる学習者の誤答を含む文を示す。c では、イメージキャプションニングによって生成された画像から抽出されたキャプションを表す。次に d では、抽出した文と、元となる文の一致度を示す。ここでは、画像に含まれる情報を自然言語化する手法とし

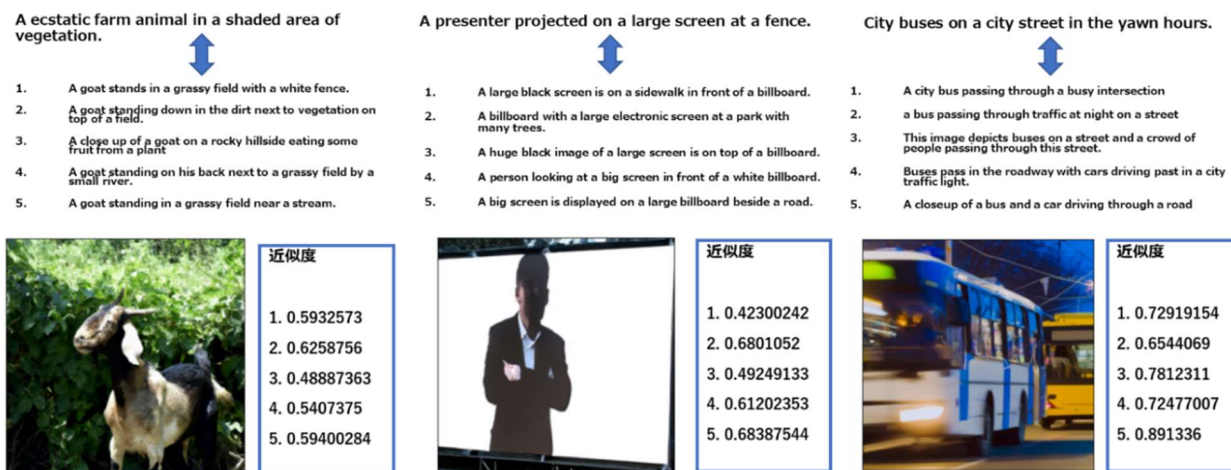


図 9 特徴消失した生成画像に於けるコンテキストの類似度

て、文脈を加味した単語の分散表現を得ることが可能である。BERT⁽¹¹⁾の拡張として、文の分散表現を得ることが可能である。Sentence BERT⁽¹²⁾を用いることで、文の一致度についてコサイン類似度を基に推定する。この時、任意の閾値を超えたとき、生成された画像は、入力に対する文の特徴を有していると評価が可能である。図 9 では、3.4 で言及した特徴消失が発生した画像に対する、生成画像と抽出した文の近似度を示す。ここでは、図 8 で示した生成画像と比較し、特徴消失を有する画像は一部のキャプションに於いて、高い数値を示す。つまり、学習対象語の特徴を有していないが、元の文のコンテキストは類似していると捉えられる。

4.2.2 誤り概念の類似度

学習対象語の一致度を測定する上で、先述した BERT のツールの一つである、BERT Tokenizer を利用し、前後の文脈を加味した対象単語の分散表現を得る方法を検討する。図 10 では、学習者の誤答を含む「Two sheep and a (tram) stand next to a fence in the yard」の文中の、誤答である学習対象語「tram」について、BERT Tokenizer により前後の文脈を加味

した単語の分散表現を取得し、その後、1 のプロセスで得た文に含まれる単語全てに対して、一致度を推定する。この時、最も一致度が高い数値が任意の閾値を超えた場合に、画像が学習対象語の特徴を含意しているか否かの定量的な評価が可能であると捉えられる。

4.2.3 評価方法の考察

現在はコンテキストの類似度に於いて、生成画像より 5 つのキャプションを抽出し、それぞれの文に対する類似度を基の画像と比較し評価した。一方、図 8 の例では、5 番目の文が最も類似度が高いとされるが、一方、学習対象語となる「tram」の特徴を有するという観点では、1 番目の生成文書に「track」という単語を有する点からも、誤りの類似度という観点では、1 番目の文書のほうが類似度が高いと言える。つまり抽出した各文書に対する類似度の推定では、特徴の見落としが発生する可能性を有する為、抽出したキャプションすべての特徴と、元となる文の特徴の比較という方法も考えられる。又、この観点からも 5 つのキャプションで十分であるかは今後検討の必要があり、どの程度のキャプションの生成が必要であるか今後の検討が必要である。

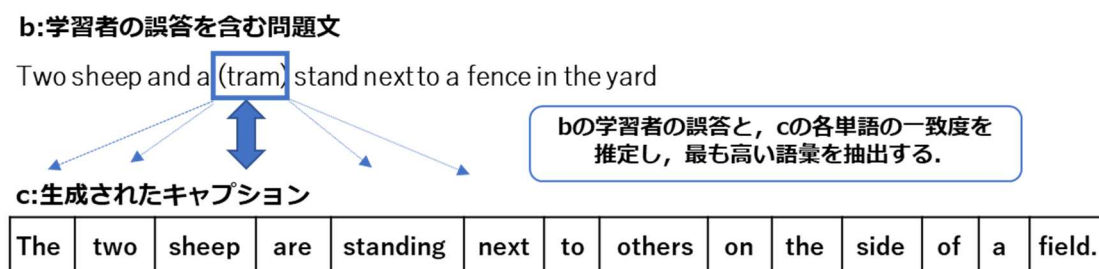


図 10 誤り概念の類似度の推定

5. まとめ

提案手法では、誤答時の画像変化が繰り返しの誤答を防ぐうえで有効である事を示した。又、画像に含まれる、印象、想起性、の二つの評価軸が生成画像の示唆性の指標として必要であると定義した。実験結果と考察より、想起性より印象の値が繰り返しの誤答を避ける上で有効であったが、想起性が欠如する特徴消失問題は、誤答を理解するに足る情報が不足する問題を孕む。そこで、誤答時の画像変化に依る語彙学習支援に於ける有効な画像を評価する上で、印象と想起性に於ける定量的評価指標の必要性に触れ、特に想起性に於ける定量的評価指標について検討した。ここでは、それぞれ、1. コンテキストの類似度、2. 誤り概念の類似度と定義し、方法論について言及した。一方、生成画像に含まれる想起性の評価指標として、1, 2 の数値に於ける閾値を定義する必要がある。今後は、1, 2 の数値と学習効果の関係性から、それぞれの数値の重みを定義する事で、生成画像の想起性に対する評価指標を策定する必要がある。又、効果的な学習という観点から、印象に対する評価指標の策定も必要となることから、今後検討していきたい。誤答時の生成画像に於ける定量的な評価指標の策定は、今後画像生成を活用した語彙学習支援システムによる基準となり、多義語、類義語といった様々な特徴を有する語彙の可視化の指標となることを期待している。

謝辞

本研究は、JST 次世代研究者挑戦的研究プログラム JPMJSP2102 の支援を受けたものです。

参考文献

- (1) Krashen, S. D. “Principles and practice in second language acquisition”, 1st ed. Pergamon, Oxford; New York, (1982)
- (2) Hasnine, M. N, Flanagan, B., Akcapinar, G., et al. “Vocabulary Learning Support System Based on Automatic Image Captioning Technology,” in Distributed, Ambient and Pervasive Interactions, vol.11587, N. Streitz and S. Konomi, eds. Springer International Publishing, Cham, pp.346–358, (2019)
- (3) Vigil N. and Oller, J. “Rule fossilization: A tentative

model,” Language Learning, vol.26, pp.281–295, Oct. (2006)

- (4) Sugita, K., Gu, W., and Hasegawa, S., “Error-Driven Learning: Development of Vocabulary Learning Support System that Utilizes Suggestibility of Error by Image Generation,” IIAI Letters on Informatics and Interdisciplinary Research, vol.3, p.1, (2023)
- (5) Lin, T.-Y., Maire, M. Belongie, S. et al. “Microsoft COCO: Common Objects in Context,” arXiv, (20-Feb-2015)
- (6) Ramesh, A., Dhariwal, P. Nichol, A., et al. “Hierarchical Text-Conditional Image Generation with CLIP Latents,” arXiv, (12-Apr-2022)
- (7) Sugita, K., Gu, W., Ota, K., and Hasegawa, S., “L-VEIGe : Development of L2 Vocabulary Learning Support System with Suggestibility of Error by Image Generation’ IEICE Trans. Inf. & Syst (under review)
- (8) Hessel, J., Holtzman, A., Forbes, M. et al. “CLIPScore: A Reference-free Evaluation Metric for Image Captioning,” Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Online and Punta Cana, Dominican Republic, pp.7514–7528, (2021)
- (9) Radford, A., Kim, J. W., Hallacy, C. et al. “Learning Transferable Visual Models From Natural Language Supervision,” arXiv, (26-Feb-2021)
- (10) Xu, K., Ba, J.,Kiros, R., et al. “Show, Attend and Tell: Neural Image Caption Generation with Visual Attention,” arXiv, (19-Apr-2016)
- (11) Devlin, J., Chang, M.-W., Lee, K., et al, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” arXiv, (24-May-2019)
- (12) Reimers, N. and Gurevych, I., “Sentence-BERT: Sentence Embeddings using Siamese BERT Networks,” (arXiv, 27-Aug-2019)

英語教育支援システム Zoroe を用いた

ディクテーション分析に基づくリスニング指導

照井雅子^{*1}, 溝渕昭二^{*1}

^{*1} 近畿大学

Developing listening skills based on dictation analysis using an English teaching support system Zoroe

Masako Terui^{*1}, Shoji Mizobuchi^{*1}

^{*1} Kindai University

The first author reports on the effectiveness of the system in developing listening comprehension of 15 first-year university students by analyzing their own English dictation results using the alignment function of Zoroe, a teaching support system developed by the second author. In dictation, learners are asked to write down what a speaker said as correctly as possible. This enables learners to focus on specific language words and/or phrases, grammatical errors as well as syntactic mistakes. Yet, it takes a lot of time for teachers to check the dictation results of their students in the conventional way, where dictation results are compared to an original sentence word by word. Therefore, this English teaching support system Zoroe was developed to allow learners of English to find for themselves "where" and "why" they were unable to catch words. Teachers do not have to take time to check students' dictation results word by word and students do not have to repeat repetitive dictation tasks. Learners' autonomy is one of the most innovative and impressive learning effects. Through this system, students can analyze the dictation results of not only themselves but also other learners in class by seeing aligned dictation results at a glance.

キーワード: 英語教育支援システム, ディクテーション, リスニング指導, 学習者自律, 外国語教育

1. はじめに

1.1 ディクテーション

ディクテーションとは、学習者がリスニングを行う際に話者の発話（一般的には学習用音源の音声）を聴き取ったとおりに記述し、元の発話と比較することである。外国語教育において、リスニング力を測る手段として、また、ディクテーションした結果を元の発話と正誤を確認する際の気づきによってリスニング力を向上させる手段として、長く用いられている。聴き取ったものを一言一句正確に記述することが求められ、何が聴き取れて何が聴き取れなかったかを知ることができるため、学習上大きな利点がある。

学習効果の面からは、単語や語句に留まらず、文法の間違いや文の構成にも注意を向けられるセンテンス単位のディクテーションを行うことが望ましい。

ただし、ディクテーション自体ではさほどではないが、正誤を確認する所謂「答え合わせ」に時間がかかる。従来、学習者のディクテーション結果と元の発話のスクリプト原稿を突き合わせ、単語レベルで比較して正誤を確認しているが、学習者自らが行うにしろ教員が行うにしろ、その労力が大きい。特にセンテンス単位のディクテーションには時間的制約と正誤確認の労力が大きく、ディクテーションがリスニング力を向上させる一定の効果が認められつつも⁽¹⁾⁽²⁾⁽³⁾⁽⁴⁾、授業内

の課題として導入されることが敬遠されがちである。

1.2 英語教育支援システム Zoroe

前節のとおり、ディクテーション指導の際に、複数の学習者が記述した複数の（多くの場合、多量の）英文から誤りを見出すのに必要な時間と労力が大きいことが問題である。その負担を軽減することを目的として、自然言語処理を専門とする第二著者によって英語教育支援システム Zoroe が開発された。システムが持つ英文アラインメント機能によって、複数の英文を単語単位で揃えて一覧表示できることが特長で、システムの名前「Zoroe」の由来ともなっている。

教員や学習者自らがそのアラインメント結果を分析することで、ディクテーションで正解に至らなかった誤りそのものや、誤りの傾向の把握が容易になる。

2. 研究の目的

ディクテーションにおける難語とは、主に内容語すなわち未知語である。語彙力がディクテーション結果に大きく影響を与えており、その内容語が語尾変化を伴った場合に、特にその語彙の書き取りに困難を覚える傾向が見られる^⑤。つまり、日本語母語話者である学習者が、異なる音声体系を持つ英語の音声情報を正しく書き取ることは極めて難しい。

そこで、比較的短い単文として、TOEIC® Listening & Reading の写真描写の音声を手書きでディクテーションさせ、その結果を英語教育支援システム Zoroe のアラインメント機能を利用して分析させることにより、音声に関して「どこで」「なぜ」聞き取りが困難であったのかを学習者自身が客観的に分析できることを確認する。

3. 調査

3.1 調査対象者

調査対象者（以下「学習者」）は、私立大学理工学部 1 年生で、必修科目である「英語演習 1」の受講生 15 名である。学習者は、1 学年を習熟度別に 4 クラスに分けた中で最も習熟度が低いとされるクラスに在籍するが、出席率はほぼ 100% で、課題に真面目に取り組

んでいる。クラス分けには入学時に受験した初・中級者向けの TOEIC Bridge® Listening & Reading Tests スコアを利用した。TOEIC® Listening & Reading のスコアに換算すると、クラスの平均は 256 点で、中央値は 263 点であった。基礎力をつけることが課題となる初級者レベルと言える。

基礎力が必要な学習者だからこそ、自身が「どこで」「なぜ」聞き取りが困難であったのかに気づくことが重要である。

3.2 調査手順

3.2.1 調査期間

2023 年度前期に開講されている 1 年生の必修科目「英語演習 1」の課題として 5 月と 6 月の授業内に行った。

3.2.2 ディクテーション素材

学習者が在籍する学部のオリジナルテキスト^⑥に含まれる TOEIC® Listening & Reading 対策のオリジナル練習問題の一部を利用した。

3.2.3 Zoroe による分析用データセット

練習問題として 10 枚の写真それぞれを描写する英文選択肢が 4 つ読み上げられ、学習者 15 名それぞれが紙に手書きした。各自 40 文のディクテーションを行い、おおよそ 600 文のデータセットが得られた。

3.2.4 ディクテーション課題の例

以下の図 1 にディクテーション課題の例を示す。読み上げられた文（正解）は下線で示した。



図 1 ディクテーション課題の例

ディクテーションさせた文は以下のとおりである。

- (A) The dog has its head on the pillow.
- (B) The dog is resting beside the woman.
- (C) The woman is sweeping the bedroom.
- (D) The woman is taking a nap on the floor.

これらの文をディクテーションさせることで、「bed」や「sleep」に関連する「pillow」「take a nap」といった表現や、「sleep」に発音や綴りが似ている「sweep」が正しく聞き取れたかを確認できる。

4. 結果

4.1 Zoroe



図2 Zoroe のログイン画面

The	dog	has	its	head	on	the	pillow
The	dog	-	is	handed	on	the	pillow
The	dog	-	is	head	-	-	-
The	dog	has	-	-	-	-	pilor
The	dog	has	-	head	on	the	pillow
The	dog	has	-	hed	a	-	ピロー
The	dog	has	his	head	on	the	pillow
the	dog	has	it	-	-	-	below
The	dog	has	it	head	on	the	pelow
The	dog	has	it	head	on	the	pillow
The	dog	has	it's	head	on	the	pillow
The	dog	has	it's	hed	on	the	pirow
The	dog	has	its	head	on	the	pillow
The	dog	has	its	head	on	the	pilo
The	dog	has	its	head	on	the	pillow
The	dog	has	its	hed	on	the	pillow

図3 Zoroe のアラインメント機能

4.2 Zoroe のアラインメント機能についての学習者のメリット

Zoroe のアラインメント機能は、自身のディクテーション結果の正誤の確認だけでなく、同じクラスの他の学習者のディクテーション結果が一覧できるため、学習者に高く評価された。主な意見を示す。

- ・みんなの答えの表がすごくわかりやすいと思いました！
- ・表で見て分かったことは、できるところはみんなできている。できていないところは同じような間違いをしていた。

・表で見ると、どこの単語が聞き取りにくいかが一目でわかるので良い。同じような間違いをしている人が多いということを知ることができたので良かったと思う。

その上で、教員が望む「学習者自律」の効果を示す意見も見られた。特に、日本語話者である学習者にとって難しいとされる、子音と母音、単語と単語が連結するなどして音が変わるリエゾン (liaison) や、文の構造そのものや冠詞の難しさについて自ら気づけている点は Zoroe のアラインメント機能が奏功していると言える。主な意見を示す。

- ・音がつながっているのが聞こえていない人が多い。単語どうしがつながっているように聞こえたりして、正確に聞こえていないので、そのことを考えながらリスニングする。
- ・主語は書いている人が多いが、動詞や目的語は書けていないことが多い。

・主語をしっかりと聞き取れると、文の構造も想像しやすい。

- ・次にどんな単語が出てくるか予想する。
- ・a と the の違いに気を付ける。違いを聞き取るにはどうしたら良いか。

さらに、自ら、語彙習得の重要性にも気づけている。

- ・知らない熟語は聞けても書きにくい。意味がわからないから。
- ・pile of が piled に聞こえた人が多い。pile を知っていたら、of がくるのが分かったと思う。

4.3 Zoroe のアラインメント機能についての教員のメリット

従来のように、学習者のディクテーション結果を元の発話と単語レベルで正誤を確認するだけでは全体的な傾向を把握しにくく、教員のある種の勘に基いた指導になってしまう可能性がある。しかし、Zoroe のアラインメント機能を用いてクラス全体の結果を一覧することにより、例えば L と R の発音の聞き取りが難しかったために pillow を pirow と誤った綴りを書いてしまったり、pillow という語彙が定着していない可能性があることに教員が気づき、L と R の聞き分けを集中的に指導したり、語彙を強化する課題を課したりでき

る.

5. まとめ

5.1 Zoroe のアラインメント機能の効果

前章で述べたように、学習者はディクテーションした結果を元の発話と単語レベルで正誤を確認するのではなく、Zoroe のアラインメント機能を用いて学習者全員のディクテーション結果を一覧できることにより多くの気づきを得られている。

5.2 今後の課題

今回は習熟度が低い学習者が Zoroe のアラインメント機能を用いて分析を行い、自ら誤りや今後の学習で気をつけることに気づけたことがわかった。今後は習熟度の異なる学習者、特に習熟度が高い学習者を対象として同じ調査を行い、Zoroe のアラインメント機能はどういった学習者により効果があるのかを確かめたい。

参 考 文 献

- (1) Oller, J. W., Jr.: “Dictation as a device for teaching foreign language proficiency”, *ELT Journal*, 25(3), pp. 254-259 (1971)
- (2) Morris, S.: “Dictation: A technique in need of reappraisal”, *ELT Journal*, 37(2), pp.121-126 (1983)
- (3) Weir. C.: “Understanding and developing language test”, Prentice Hall, USA (1993)
- (4) Hughes, A.: “Testing for language teachers (3rd ed.)”, Cambridge University Press, UK (2003).
- (5) 松岡博信, 平本哲嗣: “大学生の英語リスニングおよびディクテーションにおける困難点: TOEIC 問題演習結果の分析”, 中国地区英語教育学会研究紀要, 第 36 号, pp.79-88 (2006)
- (6) 近畿大学理工学部英語教育研究会編: “First Year in RIKO 2023”, 成美堂, 東京 (2023)

初修外国語向け HyFlex 型授業デザインの最適化

—日本の三大学における実践報告—

杉江聡子^{*1}, 王松^{*2}, 曲明^{*3}, 砂岡和子^{*4}

^{*1} 札幌国際大学 ^{*2} 関西外国語大学短期大学部 ^{*3} 室蘭工業大学 ^{*4} 早稲田大学

Optimal HyFlex Course Design for Basic Foreign Languages Class

—A Report of the Three Japanese Universities—

SUGIE Satoko^{*1}, WANG Song^{*2}, QU Ming^{*3}, SUNAOKA Kazuko^{*4}

^{*1} Sapporo International University ^{*2} Kansai Gaidai University

^{*3} Muroran Institute of Technology ^{*4} Waseda University

日本の教育機関での遠隔授業は、新型コロナの大流行以来、試行錯誤しながら進展した。このうち、米国で既に普及している HyFlex 型授業は、日本では会議や渡日前学生向けの運用が多く、授業実践は多いとは言えない。本研究は HyFlex 型授業の要点を総括し、三大学での初修中国語 HyFlex 型授業を事例に実態と課題を整理し、改善策を検討した。大学授業のステークホルダー別に評価した結果、組織は継続してハイブリッド型授業に適したインフラ環境の整備構築と通信機器運用支援制度を改善すること、教員は両モードの活動やインタラクションを意識した教材作成やタスク設計が不可欠であることが明らかになった。北米と日本では教学環境や学修者のモチベーションが異なり、日本の学生は遠隔モードでの協同学習が不活発となるため、授業設計や教材準備の段階で HyFlex 型授業に適したインストラクショナルデザインが必要なことが確認された。

キーワード: ブレンディッド学習, 遠隔教育, HyFlex, 語学教育, インストラクショナルデザイン

1. はじめに

2019 年冬以来、新型コロナのパンデミックで世界の教学現場は激変した。日本の高等教育機関では半強制的に遠隔授業やオンライン課題を実施する形となった。北米では以前からブレンド型学習やハイブリッド型授業が普及しており、対面と遠隔の異なる 2 つのモード（オンデマンドを加え 3 モードの場合もある）を同期で運営する HyFlex 型授業 (Hybrid-Flexible Course) が 2000 年代から開発されていた (Means, Toyama, Murphy, Bakia & Jones, 2010)。ブレンド型学習は高等教育の正当で優れた形態として確立していたが、教師主導型が主流で、学習者主導の教室対面と遠隔の両

モード運営の知見不足が指摘され、Hybrid と Flexible を融合した HyFlex 型授業が登場した。HyFlex とは、学生が授業を対面で受けるか、同期オンラインで受けるか（+オンデマンド受講）を選択できるクラスセッションと定義され (SFSU Academic Senate Policy S16-264), ①少なくとも 2 つ（対面+最低 1 つのオンライン）の学習経路を完全に統合する, ②これらの学習経路の中で、学生の継続的な選択（柔軟性）を支援する, という 2 点を意図的に組み込んだデザインである。HyFlex 型授業は入学者募集と在学維持という大学運営面にも利すると考えられ、遠隔地の学生に利便性や受講の選択肢を増やし、突発的事象、天変地異、緊急事態における教育と大学運営継続を目的に定

着した。

日本では新型コロナの感染予防対策として、対面活動が制限され、非対面型教育方式が模索される中で、遠隔授業が登場し、各教科で多様な実践が展開された。中国語教育でも、コロナ以前から ICT 活用を推進してきた教育実践の知見が注目され、オンラインセミナーや研究報告の機会も増えた（清原, 2022）。しかし、情報工学関連分野と比較すると、外国語教育の実践報告や知見の蓄積は多いとは言えない。国立情報学研究所ではパンデミック初期から情報学研究者と教育者が連携し、コロナ禍での教学の課題を概観し、次世代の日本の教育を議論するオンラインシンポジウムを開催してきたが、大規模校による報告が大半で、外国語教育に関する発表は国内外を含め僅かである（国立情報学研究所, 2020-現在）。本研究では、HyFlex 型授業設計の理念（Beatty, 2019）に基づき教学環境の要点を整理した上で、日本の初修中国語に適した HyFlex 型授業デザインを検討する。日本の三大学の中国語授業に基づき、教学の中心的主体である教員と学習者、および教育観・学習観の特徴に照らして、北米発の HyFlex モデルを日本型に最適化する道を探る。

2. HyFlex 型授業設計の基本的価値と原則

2.1 HyFlex の原則

HyFlex 型授業設計の開発は、2006 年初頭から、学習者の選択（Learner Choice）、同等性（Equivalency）、再利用性（Reusability）、アクセシビリティ（Accessibility）という 4 つの基本的価値と原則に基づき進展した。教授設計は“drivetrain”と呼ばれ、価値観から始まり、そこから教育・指導の目標が導き出され、それに向けて全体の教授方略が選択され、文脈を考慮した特定の教育活動によって実践される。「学習者の選択」とは、選択可能で有意義な受講モードを提供し、学生が期間やトピックごとに受講モードを選べるようにすることである。学習者自身による選択の柔軟性がなければ、従来のハイブリッド型授業と変わらない。「同等性」とは、全ての受講モードで同等の学習成果が得られるよう、学習体験を提供する原則で、HyFlex の最大の課題である。「再利用性」とは、各モードでの学習活動からの成果物を、更に学習コンテン

ツとして再利用可能にすることである。教師が準備するマルチメディア教材に限らず、学生活動により生成したチャット、非同期のディスカッション記録、ファイル投稿、ピアレビュー記録等は、対面学生の良き学習支援コンテンツとなり、オンライン学生にとっても有益なレビュー資料となる。「アクセシビリティ」とは、全ての受講モードで、技術スキルと公平なアクセスを学生に提供することを指す。受講モードを適切に選択するには技術的なスキルが求められ、学生と教師にリソースとトレーニングを提供することが義務付けられる。また、全ての教材と活動に全学生がアクセスできるよう、必要な施策を講じなければならない。

2.2 HyFlex 型授業のデザインと評価

Beatty（2019）は HyFlex 型授業の導入に、①メリットとコストの比較評価→②学生が期待する学習成果の分析、確認または修正→③学生の学習活動（授業内活動とインタラクション）計画→④学習成果評価のための準備→⑤期待に対するリターンの評価、という 5 つのステップを推奨している。2007 年以降 13 論文のレビューによれば、HyFlex 型授業研究の関心の焦点は、授業参加度や成績評価（教育効果の有無）から、学生のパフォーマンスに影響する要因や学習共同体としてのダイナミクスに、更には学修満足度の大規模調査へと変化している（Beatty 2019）。受講形態は、原則、対面、同期遠隔、非同期遠隔（所謂オンデマンド）の 3 タイプを保障すべきだが、実際は受講形態や授業過程、活動、使用ツールなどが異なる“Modified version”が多数存在する。HyFlex 型授業に関する研究は、教育効果の一般化と妥当性の検証を目指す量的研究が中心で、学習モチベーションがパフォーマンスの強力な予測因子であることが指摘される（Helms 2014, He, Gajski, Farkas & Warschauer 2014）。また、HyFlex モデル導入により、物理的・仮想的な学習空間（例：LMS）の拡張可能性や、教師と学生・学生同士の新たなインタラクションと連帯が生まれ、学生のエンゲージメントを高める効果を示唆する研究がある（He et al. 2014, Miller & Griffiths 2013）。HyFlex 型授業デザインの研究は北米と豪州に多く、アジアが少ないのは、地域により教育文化の差異が大きいためと推測できる。

2.3 HyFlex デザインのメリットとデメリット

北米の高等教育ではハイブリッド形態による学修意義について、10年以上の研究蓄積がある。遠隔と対面授業を比較した45論文のメタ分析によれば(Beatty, 2019), 遠隔受講学生は対面授講学生より平均してやや成績が良く、中でも「オンライン+対面の組み合わせ」と「完全対面」の成績差が大きい¹とされる(Means, Toyama, Murphy, Bakia, & Jones, 2010)。ブレンド型もしくはハイブリッド型授業が、オンラインのみ・対面のみより成績が良い理由は、学修時間の増加に加え、教育資源と学習活動が増えることによる。従来の対面やオンライン授業は、教育活動が固定され、学生側に受講形式や出席選択の自由がなく、たとえ教員側が授業デザインに工夫を凝らしても、スケジュールが重なれば、学生は授業を欠席する他なかった。HyFlexは対面とオンラインの両モードの利点を組み合わせたデザインで、学生の授業参加を保障し、学修成果の最大化を目標とする。ただしモードの組み合わせかたによって、最適な学習ゴールや教育方略等が異なるため、

単純に学生の成績だけで形態の優劣を判断することはできない。どのような状況でHyFlexアプローチ導入がコストに見合うのか、自身の大学・学科・コースの条件との適合性を入念に検討すべきである。以下、日本の三大学でのHyFlex型授業実践について、授業環境や条件に焦点を当てて比較分析する。

3. 三大学 HyFlex 中国語授業の自己評価

3.1 三大学における HyFlex 中国語授業の設計

2021年度秋学期、日本の3つの大学(A大、B大、C大)で初修中国語のHyFlex型授業を行った。各校の授業環境と諸条件を表1に示す。3校に共通の要素は以下諸点である。①専任教員または非常勤講師とのペア担当、②Zoomによる教室からの遠隔講義配信、③ゼロ基礎学生、通年授業(春学期に中国語Ⅰ、秋学期にⅡを履修)④選択必修科目(A大・C大)または自由選択(B大)、⑤教科書準拠の語彙・文法中心学習。

表 1 三大学の HyFlex 中国語授業の環境と条件

	A 大 (国立)	B 大 (私立短大)	C 大 (私立)
(1) 教学主体			
教員・教歴	専任・中国語ネイティブ, 約 21 年	専任・中国語ネイティブ, 約 20 年	専任・日本語ネイティブ, 約 20 年
クラス編成	50 余名 × 3 クラス	約 20 名 × 1 クラス	5 ~ 10 名 × 2 クラス
授業数	週 1 コマ (90 分)	週 2 コマ (180 分)	週 1 コマ (90 分)
学習者	日本人・工学系 1 年	日本人・文系 (学年混合)	留学生を含む, 複数文系・学年混合
(2) 大学・授業実施環境			
実施期間	2021 年 10 月上旬～翌年 1 月下旬	2021 年 9 月中旬～12 月	2021 年 9 月下旬～翌年 1 月下旬
授業形態	遠隔 2 回: HyFlex13 回	HyFlex29 回: 遠隔 1 回	HyFlex14 回: 遠隔 1 回
学生受講形態比 (平均)	対面 1 : 遠隔 5	対面 1 : 遠隔 2	対面 1 : 遠隔 2
通信 (大学)	学内有線 LAN+Wi-Fi	学内 Wi-Fi	学内有線 LAN+Wi-Fi
通信 (学生)	自宅 Wi-Fi	自宅 / 外部 Wi-Fi	自宅 Wi-Fi
教室	PC 教室 (各自 PC+ネット接続)	通常教室+スクリーン	PC 教室 (各自 PC+ネット接続)
遠隔ツール	Zoom, Teams, One drive	Zoom, 広画角 WEB カメラ(logicool MeetUP)	Zoom, 広画角 WEB カメラ(logicool MeetUP)
教員端末	ノート PC	ノート PC	ノート PC (含 TA), タブレット
学生端末	PC (全員)	PC (少数), スマホ (多数)	PC (全員)
LMS	Moodle 型 LMS	LMS (Blackboard)	HTML 型 LMS
ICT 支援	有 (常時)	有 (必要時に依頼)	有 (必要時に依頼)
TA	無	無	2 名 (中国人院生)
(3) 授業運営形態			
教授法・授業形式	反転授業を併用	シラバスベース	シラバスベース+課内外グループ活動
教材	電子版自編教科書+授業録画	市販教科書+スライド+副教材	市販教科書+スライド+副教材
試験	期末+毎週小テスト (オンライン)	期末+毎週小テスト (オンライン)	期末 (対面), 小テストなし

¹ 実験群と統制群の差を標準偏差で割った測定値。

3.2 HyFlex 中国語授業に対する三大学教員の評価

HyFlex 型授業の達成状況につき、三大学の担当教員が各自の実践知に基づき、総合評価と項目別評価を行った（表 2, 3）。各指標の項目作成にあたり、Beatty（2019）と浦田（2022）を参照し、HyFlex 型授業のメリットとデメリット（課題・コスト）に大別して下位項目を設け、更に大学授業のステークホルダーとして「組織」「教員」「学生」に三分類し、授業担当教師の視点から評価した（1 から 3 の 3 段階評価）。「デメリット（課題・コスト）」は評点が高いほどその項目のデメリット値が大きいことを示す。教員の立場からは判断できない項目は N とした。いずれも主観的な評価で、教員間の判断尺度の揺れも未統制なため、以下引用数値は相対的参照データである。

表 2 HyFlex 理念の付加価値に基づく総合評価

評価項目	平均値
学習経路の多様化	2.3
学習選択肢の合理的配慮	3.0
協働学習（アクティブラーニング）	1.7
協働学習（自律学習）	2.0
教学リソースの再利用	3.0
公正性	2.3

(1) 総合評価

表 2 のとおり、教員は HyFlex の理念を高く評価している（三大学の合計平均 2.4）。特に、「学習選択肢の合理的配慮」と「教学リソースの再利用」の原則への共感が高い（平均 3.0）。「学習経路の多様化」や「公正性」理念も次いで高い（平均 2.3）。一方、HyFlex の「協働学習」への評価はやや低く、「アクティブラーニング」と「自律学習」の評価は、A 大を除く 2 大学で 1.0 から 2.0（三大学平均は 1.7 と 2.0）で、課題が多いと判断している。このように、HyFlex の付加価値のうち、学習機会や選択肢の公正性に関してはポジティブな評価であるのに対し、協働学習効果にはネガティブな評価となった理由は、中国語科目の性質に依拠すると考えられる。語学系科目、特に初修者向けの外国語は、講義型授業と異なり、教員と学生および学生間の実技を伴うインタラクションが不可欠である。三大学の教員はともに、コロナ禍以前から学習者主体のア

クティブラーニングを重視していた。しかし HyFlex 型授業に変わり、オンライン受講学生が予想以上に増え、かつほぼ全員のカメラが OFF であった（A 大、B 大）。五感情報が欠落した同期遠隔モードでは、対面と同等の条件で授業中の対話活動を展開できない。ペア／グループワーク活動は更に困難である（A 大、B 大）。教員は学生の理解度把握や落ちこぼれ防止にエネルギーを費やし、頻繁に学生へ質問し、学習状況を確認する必要に迫られた。結果的に発音の反復練習や文型のパターン練習など、授業内容を単純化、単調化せざるを得なかった（A 大、B 大）。

HyFlex の「学習選択肢の合理的配慮」により、学生はオンラインでも受講可能となり、学習経路の選択肢は増えた。更に「教学リソースの再利用」により、学生の練習時間が増加した。コロナ以前から反転授業を実施していた A 大の学生は、一定程度オンライン課題に取り組む習慣が定着しており、教学リソースをうまく再利用できた。そのため HyFlex 型授業期間で単位を落とした学生は少なく、期末試験の成績も優秀であった。B 大は 2020 年度の教学リソースを再編集して用いた。成績は口頭試験では対面とほぼ変わらず、これは教員が遠隔でも学生の発音や暗唱・朗読確認に注力した結果であろう。オンライン実施の小テストと中間・期末試験は、前年度対面期より高得点であった。ただし、学生の回答には他資料参照の可能性が否定できず、個人の知識や能力の習得度を測る試験成績と同一視できない。C 大はコロナ以前に LMS が導入されておらず、学生に反転授業の経験もなかったため、教員が準備・指示したオンライン課題に取り組むに留まった²。また、三大学共通して、授業内活動の内容は単調で、教員と学生のインタラクションによる「協働学習」が起これにくかった。なぜ初修者向け HyFlex 型の外国語科目で、学生の「アクティブラーニング」や「自律学習」を促進できないのか、次節で詳述する。大学組織と関連する課題については、3.2 および 4 章で述べる。

(2) ステークホルダーごとの評価

(a) 教員に関する項目

HyFlex 型授業配信は複数モードにわたる管理コス

² 成績の変化はコロナ前の関連データがないため不明。

トの高さが課題である。今回の評価では教員の「情報機器操作スキル、学生管理、フィードバック、トラブル支援」への評価は（三大学平均 2.7. 以下同様に平均値を記載）と高いことから、教員が複雑なモードに適宜対処したことがうかがえる。ただし、教員には相応の「認知的負荷」（2.0）もかかっている。

また、三大学の教員は、「教学リソースの再利用」、「対面授業の代替手段」（ともに 3.0）、および「教学リソースの共有」、「連携指導促進（課題遂行）」（ともに 2.3）のメリットを高く評価している。これは前掲の「HyFlex 理念の付加価値：教学リソースの再利用」の評価が高いことと一致する。同期遠隔モードを含む授業配信では、教材等の電子データの送受信や共有が不可欠であり、課題遂行に必要な教員/学生⇄教材間の連携が促進されたという担当教員の認識が読み取れる。三大学ともに、従来の紙版教科書以外に HyFlex 型授業向けの説明資料や練習問題（PowerPoint スライド等）を作成し、スクリーン提示や画面共有しながら、対面の学生にはプリント配布する等の対応をした。これらの資料は LMS へ保存し、課外でもアクセス可能な状態にして、「教学リソースの再利用」を図っている（表 1）。既述のように、A 大は以前から Moodle 型 LMS でオンライン教材と課題を用いた反転授業を実施しており、「教学資源再利用」のノウハウを活かし、HyFlex 型授業への円滑な切り替えが実現した。そのため、A 大の各項評価は、おしなべて他校より高い（評価合計で A 大 66>C 大 60>B 大 46）。

C 大は少人数クラスと TA 配置に加え、教員の授業デザインが HyFlex 型授業でのアクティブラーニングを可能にしている。Zoom のブレイクアウト機能を使い、教室側と遠隔側でそれぞれグループを組み、中国人 TA 主導による発音指導や会話練習といった自由会話の時間も確保した。B 大もシラバス準拠の規定が強く、調整の余地が少ない。学生全員に口頭発表等の機会を確保したが、カメラ OFF で学生の表情が掴めず、対面同様の効果は得られなかった。そのため、課外での予復習を強化し、教員は随時フィードバックを行った。授業運営の負荷は増したが、学生の学習量は増える結果となった。

既述のように、教員から見た連携指導に関する効果のうち「知識習得」と「(技能習得)」の促進作用に対

しては低い評価（ともに 1.3）となった。A 大は反転授業に小テストを組み込み、毎週予習として課すことで授業との連携を図ったが、アクティブラーニングの効果は予想を下回った。

（b）学生に関する項目

「学生の受講教材選択肢拡大」は教員による評価が低い（1.3）。学習活動がオンライン化したことで「再利用可能」な教材は増えたが、教員が教材として適切に加工し、受講モード別に配信する知識と時間が不足していた。教材を準備したとしても、授業運営が煩雑になり、HyFlex 型授業配信に伴う教員の「認知的負荷」（2.0）が増加すると予想される。

コロナ禍で急遽始まった日本のオンライン教育は、「授業を止めない」ことを最優先する緊急措置であり、組織も教員も、遠隔教育の経験や知見の蓄積がないまま実践を迫られた。特に「遠隔インタラクティブ活動計画」（Beatty 2019 の「HyFlex 導入ステップ③」）の検討不足は、オンライン上で「学生の学習活動」が不活発に終わった原因と言える。対面に比べオンライン授業は五感情報の欠落を免れず、教員と学習者あるいは学習者間のインタラクティブが難しいため、参加者の社会的存在感や帰属感が薄れる（Piccoli et al., 2001）。Swan (2001)によれば、学習満足度は教室での教員との交流頻度と相関があり、教師やクラスメートとのインタラクティブが充分なら、学生の満足度と学習レベルが高い傾向にある。中でも「協働学習」は、対面か遠隔かを問わず、「アクティブラーニング」や「自律学習」を促進し、殊に語学授業では参加者間のインタラクティブを活性する極めて重要な要因である（Gass& Mackey 2020）。機関も教員もこれらの課題の検討が不十分なまま、形式的に遠隔や HyFlex 型授業を実施した結果、教員と学生あるいは学生間の「協同学習」が起こりにくく、単調な活動に偏りがちとなった。

（c）組織に関する項目

米国の教育機関は HyFlex Course Model による学生の「出席率」と「修了率」の向上を想定するが、本調査では三大学とも低い評価（1.7）となった。受講コストに敏感な欧米の学生に対し、日本の学生はもとより授業に真面目に出席し、修了率も高い。日本の大学組織は、HyFlex 導入による入学者数の増加や修了率の向上に期待しておらず、米国の大学が、革新的授業

モデルの導入により、「履修者数制限枠緩和」や「受講者数増加」を高めようとする pragmatism を日本に直接適用することは現実的ではない（表 2 で当該項目は N）。受講コストの意識が希薄な代償として、日本の学生は「時間管理能力」「学習自己管理スキル」「情報機器操作スキル」が不足しており、A 大を除き教員による評価は非常に低い（1 から 2）。受講者の自律性と積極性を前提に HyFlex 型授業を実施することの難しさが示唆される。

表 3 HyFlex 型授業設計の目標と 3 大学の達成状況

デメリット（課題・コスト）		A	B	C	平均
教員	複数モードの授業準備	3	2	2	2.3
	情報機器操作スキル	2	3	3	2.7
	複数モードの学生管理	3	2	3	2.7
	複数モードのフィードバック	2	3	3	2.7
	複数モードのトラブル支援	3	2	3	2.7
	複数モード対応への認知的負荷	2	2	2	2.0
学生	学習法調整能力	3	1	2	2.0
	時間管理能力	3	1	1	1.7
	情報機器操作スキル	3	2	2	2.3
組織	教員 FD 支援	3	1	2	2.0
	複数モードのプラットフォーム構築	3	1	3	2.3
	時間割調整・履修登録システム更新	3	1	3	2.3
メリット		A	B	C	
教員	教学リソースの共有	3	1	3	2.3
	教学リソースの再利用	3	3	3	3.0
	対面授業の代替手段	3	3	3	3.0
	連携指導促進（知識習得）	2	1	1	1.3
	連携指導促進（技能習得）	2	1	1	1.3
	連携指導促進（課題遂行）	2	2	3	2.3
学生	受講機会の拡大	2	3	3	2.7
	受講形態選択肢拡大	2	3	3	2.7
	受講教材選択肢拡大	2	1	1	1.3
	内容の選択肢拡大	N	N	N	N
	情報機器操作スキル修得	3	2	2	2.3
	学習自己管理スキル修得	3	1	2	2.0
	公正な学習機会	2	2	2	2.0
組織	履修者数制限枠緩和	N	N	N	N
	受講者数増加	N	N	N	N
	出席率向上	2	1	2	1.7
	修了率向上	2	1	2	1.7

3.3 HyFlex の課題と改善策

授業環境については、組織のインフラ環境構築・整備は、教員が教室設備や機材等を利活用する際の前提

条件である。CALL 教室や情報教育施設を使用できれば、教員と学生の端末や高速ネットワーク接続が確保され、対面と遠隔の両モードのコミュニケーション機会を十分に確保でき、LMS やオンライン課題を活用できる。新しい授業モデルと各種ツールや設備操作法の習得には、教員 FD 支援も不可欠である。CALL 教室で情報系教職員の支援を受けられる環境が整備された大学では、授業担当教員個人の技術的スキル修得や通信・機材等トラブル対応の負担は大幅に軽減される（章 2022）。支援環境がなければ、教員個人で設備機器を調達、設定する必要がある。大学が貸し出したとしても、接続設定のコストは教員個人負担となり、特に非常勤講師等は孤軍奮闘せざるを得ない。平素から専任教員との情報共有や連携が求められる（喜多，2023）。

授業運営は、教員と学生に関わる要因である。「顔出し・声出し」を避ける学生と教員がオンラインでコミュニケーションをとることは極めて難しい。個人の意志を尊重し、カメラ・マイク ON を強制することはできないが、教員と学生が語学授業でのコミュニケーションの重要性について、共通認識を持つことが根本的解決に繋がる。並行して、LMS やバーチャル空間（メタバースやアバターの活用等）での対話の可能性も探求する価値がある。

教育方針や授業形態は、組織の選択決断と密接に関わる課題である。新しい授業モデルを実施し、成功させるには、教員と学生のみならず、所属機関、ひいては行政による支援が不可欠である。2021 年以降の授業形態に関しては、文部科学省の方針もあり（文部科学省通知,2021）、各大学とも教務管理上、感染予防策を講じたうえで、可能な限り対面授業を重視する意向であった。そのため、オンデマンド授業や HyFlex を含むハイブリッド型授業の実施は積極的に推奨されなかった。教室の通信環境や設備・機材が不十分であれば、対面以外のオンライン授業を含む授業形態を適切に実施することは難しい。組織の「複数モードのプラットフォーム構築」、「時間割調整・履修登録システム更新」、「教員 FD 支援」に対する総合評価は、大学により差が大きい（1 から 3 に評価分散）。組織の支援と他の評価項目との関連性について、今後、評価尺度を細分化して分析する必要がある。

4. 初修外国語のための日本式 HyFlex 型授業設計に向けて

Raes et al. (2019) は、同期ハイブリッド型授業を設計する上で、①トレーニングとサポート、②学生と教員のコミュニケーション、③学習の活性化の3点についてのガイドラインを提案している。三大学の HyFlex 型授業実践からも、この3点に関連する課題が浮上した。①に関しては、そもそも2021年以降は教育政策の方向性に準じて対面授業へ「戻す」よう指示が出たため、HyFlex を含むハイブリッド型授業は下火となっていった。組織側も、多大なコストをかけてインフラ環境を整備することに及び腰である。一般教室は学内の Wi-Fi 回線増強が追い付かず、PC とスマホの同時接続でオンライン資源を活用できる学習の物理的前提が満たされているとは言い難い。また HyFlex のように、複数モードで受講する学生の活動を適切に管理するには、LMS システムをマルチタスク型に更新する必要がある。LMS の自動採点や、学習状態の追跡機能を使えば、教員の採点負担を軽減でき、個別学生へより細かにフィードバックできる。

教育資源の格差は教育の質に影響を与える。ウィズコロナ、ポストコロナに移行後も、また別種のパンデミックや非常事態により、対面授業が実施できない状況が起こるであろう。行政や機関は CALL 教室を始めとする教室や通信環境の改善とバージョンアップの歩みを止めてはならない。①の情報技術のトレーニングに関しては、語学授業内でのオンラインコミュニケーションや情報収集、成果物の作成など、情報スキルを使えない学生が存在する (A 大以外は評価 2)。

②に関しては、前世紀から遠隔授業を成功に導く最も重要なポイントは、心理的乖離がもたらす教員と学習者のコミュニケーション活性化であることが指摘されてきた (Moore, 1989)。今回、パンデミックで半強制的に実施した遠隔授業の体験を通じ、教師も学生もオンラインモードを取り入れることで授業がより充実することを知った (田口・鈴木 2021)。シラバスやテキストに準拠するとしても、教師による重点項目の選択や教授方略の巧拙により、必要な調整の余白を生み出し、学習意欲を引き出すことができる (牟他, 2023)。

③の学習活性化に関しては、(章, 2023) が指摘する

ように、対面とオンライン参加学生への配布教材や回答タスクの形式を参加モード別に設計することで、HyFlex 型授業での学生とコンテンツ間の相互作用を強化できる。また学生のエンゲージメントも不可欠で、短時間の活動を組み合わせて授業の流れにメリハリをつける等、学生の集中力やモチベーションの維持・向上を図り、意識的にアクティブラーニングの要素を増やす配慮が求められる (山田 2018)。

ChatGPT などマルチタスクに応答する生成型 AI ツールの急速な進化に伴い、外国語学習でも従来のような「答え」のある知識ベースの学習の意義は失われていく。今後はリアルタイムの協同学習やパフォーマンスを通じて、学習者にとっての意味や価値を創出する活動や、言語・文字情報を通じて内容の解釈や評価を行う課題を中心とする等、教学の目標、教授方略、評価指標の見直しが必須である (牟他, 2023)。テクノロジーの進化と学習者の活用スキル向上の速度を考慮すれば、ウィズコロナ、ポストコロナの教学環境は、「リアル+同期+対面式」と「バーチャルやオンデマンド+非同期+遠隔式」のハイブリッド形態が標準化するであろう。ハイブリッド型授業や HyFlex 型授業の利点と課題を正確に把握し、その遺産を対面授業の質的向上に継承・活用することが、教授者というアクターの使命と考える。

謝辞

本研究は JSPS 科研費 C (21K00773) の助成を受けた。

参考文献

- (1) Beatty, B. J.: “Hybrid-Flexible Course Design: Implementing Student-Directed Hybrid Classes”, EdTech Books (2019) <https://edtechbooks.org/hyflex/> (2023 年 3 月 28 日参照)
- (2) Bower, M., Dalgarno, B., Kennedy, G. E., et al.: “Design and implementation factors in blended synchronous learning environments: Outcomes from a cross-case analysis”, Computers & Education, Vol.86, pp.1-17 (2015)
- (3) He, W., Gajski, D., Farkas, G., et al.: “Implementing flexible hybrid instruction in an electrical

- engineering course: The best of three worlds?" Computers & Education, Vol.81, pp.59-68 (2015)
- (4) Helms, J.: "Comparing student performance in online and face-to-face delivery modalities", Journal of Asynchronous Learning Networks, Vol.18, No.1, pp. 147-160 (2014)
- (5) Means, B., Toyama, Y., Murphy, R., et al.: "Evaluation of Evidence-based Practices in Online Learning: A Meta-analysis and Review of Online-learning Studies", U.S. Department of Education, Washington, D.C (2010)
- (6) Orey, M.: "One year of online blended learning: Lessons learned", paper presented at Annual Meeting of the Eastern Educational Research Association, Sarasota, FL (2002)
- (7) Miller, J., Risser, M. and Griffiths, R.: "Student choice, instructor flexibility: Moving beyond the blended instructional model", Issues and Trends in Educational Technology, Vol.1, No.1, pp.8-24 (2013)
- (8) Raes, A., Detienne, L., Windey, I., et al.: "A systematic literature review on synchronous hybrid learning: Gaps identified", Learning Environments Research, Vol.23, pp.269-290 (2019)
- (9) Sunaoka, K. and Sugie, S.: "Remote Chinese Teaching and Learning at Japanese Universities during the COVID-19 Pandemic", In S. Liu (Eds.), Teaching the Chinese Language Remotely: Global Cases and Perspectives (pp.203-234), Palgrave Macmillan, London (2022)
- (10) 喜多一: "COVID-19 パンデミック下での大規模オンライン授業の経験と今後に向けての課題", 情報処理, Vol.64, No.2, pp.19-26 (2023)
- (11) 清原文代: "ICT を活用した中国語教育のための電子資料 及 び 教 案 の 開 発 " <https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-17K02933/> (2022 年 9 月 26 日参照)
- (12) 清原文代: "中国語オンライン授業のためのリソース", https://ja.padlet.com/kiyohara_f/e1zjgubtcgxmzco1 (2022 年 9 月 9 日参照)
- (13) 国立情報学研究所: "大学等におけるオンライン教育とデジタル変革に関するサイバーシンポジウム「教育機関 DX シンポ」", <https://www.nii.ac.jp/event/other/decs/同「教育機 DX シンポジウムアーカイブズ」> <https://edx.nii.ac.jp/> (2022 年 9 月 26 日参照)
- (14) 文部科学省: "令和 3 年度の大学等における授業の実施と新型コロナウイルス感染症への対策等に係る留意事項について (周知)" (2021) https://www.mext.go.jp/content/20210305-mxt_kouhou01-000004520-02.pdf (2023 年 3 月 26 日確認)
- (15) 田口真奈, 鈴木健雄: "オンライン授業・ハイブリッド型授業の質保証に向けてー京都大学の授業支援を事例に", 名古屋高等教育研究, No.21, pp.49-75 (2021)
- (16) 浦田悠: "ハイブリッド型授業に関する知見の整理と FD 研修の実践", 大学教育研究, No.30, pp.21-34 (2022)
- (17) 山田剛史: "大学教育の質的転換と学生エンゲージメント", 名古屋高等教育研究, No.18, 155-176 (2018)
- (18) 牟世栄, 関亨, 砂岡和子, 王怡人, 曲明, 杉江聡子, 王松: "如何教好漢語語音基礎和基本句型?" (2023) <http://www.f.waseda.jp/ksunaoka/座談会記録.pdf> [日本語訳: 砂岡和子, 向凌萱: "これだけは身に着けて欲しい発音と初級文型の教えかた" (2023) <http://www.f.waseda.jp/ksunaoka/座談会日訳.pdf>] (2023 年 3 月 30 日参照)
- (19) Moore, M.: "Three types of interaction", American Journal of Distance Education, Vol.3, No.2, pp.1-6 (1989)
- (20) Piccoli, G., Ahmad, R. and Ives, B.: "Web based virtual learning environments: A research framework and a preliminary assessment of effectiveness in basic IT skills training", MIS Quarterly, Vol. 25, pp.401-426 (2001)
- (21) Gass, S. M. and Mackey, A.: "Input, Interaction, and Output in L2 Acquisition", In B. VanPatten, G. D. Keating, S. Wulff (Ed.), Theories in Second Language Acquisition (pp.192-222), Routledge, New York (2020)
- (22) Swan, K. Virtual interaction: Design factors affecting student satisfaction and perceived learning in asynchronous online courses, Distance Education, Vol.22, No.2, pp.306-331 (2001).
- (23) 章欣: "汉语双线融合教学模式的实践路径探索", 数字化国际中文教育 (2022), pp. 251-257 (2022)

タブレット端末を用いた認知症スクリーニング検査における

自動採点機能の実装

倉持宏斗^{*1}, 小久保奈緒美^{*2}, 吉本定伸^{*1}

^{*1} 東京工業高等専門学校

^{*2} 量子科学技術研究開発機構

Implementation of automatic scoring system for tablet-based cognitive screening test for dementia

Hiroto Kuramochi^{*1}, Naomi Kokubo^{*2}, Sadanobu Yoshimoto^{*1}

^{*1} National Institute of Technology, Tokyo College,

^{*2} National Institutes for Quantum Science and Technology

今後、日本では認知症患者数のさらなる増加が見込まれており、2030年には約800万人に上ると推計されている。認知症は、個人の機能低下だけでなく高い社会経済的コストも課題となるため、国は、認知症の「予防」と「共生」を両輪とした認知症施策を推進している。筆者らは、ICTを活用し、従来のTrail Making Test: TMTを発展させ、Android端末を用いた認知機能評価用アプリケーション「User eXperience-TMT: UX-TMT」を開発し、認知症スクリーニング検査としての有用性を示した(Kokubo. et al., 2018)。本稿では、UX-TMTへ新たに実装した自動採点機能の開発について報告する。

キーワード: Android アプリケーション, 認知症, スクリーニング検査, 自動採点, フィードバック

1. はじめに

近年、日本は高齢化率の持続的高まりに伴い、認知症患者数も増加の一途をたどっており、2030年には65歳以上の高齢者の約5人に1人に達することが見込まれている⁽¹⁾。認知症は、当事者の機能やQOLを低下させるだけでなく、医療費や介護者のインフォーマルコストなど社会的コストの増大も課題である。そのため、日本は現在、認知症の「予防」と「共生」を両輪とした認知症施策を推進している⁽²⁾。

筆者らはこれまで、認知機能評価・トレーニング用Androidアプリケーション「User eXperience-Trail Making Test: UX-TMT」を開発し改良を行ってきた。UX-TMTスクリーニング検査は従来のスクリーニング検査と同様に短時間(10分以内)で行うことが可能であり、検査の完遂率も高い。また、世界で最も使用されている認知症スクリーニング検査 Mini-Mental

State Examination: MMSE-Jとの相関が高く、臨床的有用性と妥当性を確認している⁽³⁾。他にも、タブレット端末の利点を活かして反応時間や刺激提示位置、タップ座標などのデータを自動で取得することが可能である。

近年、タブレットを活用した認知症スクリーニング検査は数多く開発され、実施可能性とスクリーニング性能以外の側面にも関心が高まっている。例えば、一般的な認知機能検査は操作が容易なものでも検査と結果処理を合わせると40分程度を要するため、検査者にとって負担が小さくない。そのため、スクリーニング検査における自動採点システムの導入は、結果処理にかかる所要時間を短縮し検査者の作業量を軽減できると考えられる⁽⁴⁾。

UX-TMTの検査結果閲覧画面では、検査日時や課題ごとの正答率、反応時間、タップ位置を数値やグラフ

で表示している。また、各課題の生データを CSV ファイルとして出力している。一方、検査スコアの算出は、検査者による観察と各課題の正答率を元に検査者が行なっている。そこで本研究では、新たに UX-TMT アプリケーションによる自動採点、およびフィードバックシステムを開発し実装した。

本稿では、スコア自動採点機能とデータ保存機能、検査結果フィードバック機能の開発について報告する。

2. アプリケーション概要

2.1 検査

認知機能は注意、記憶、遂行機能、言語等の様々な要素を含む概念である。これらを包括的に評価するため、UX-TMT スクリーニング検査では4つの認知課題を1つの検査バッテリーとして実装している。また、タブレット上で検査を行うことにより、反応時間や検査所要時間、タップ位置など認知処理に係る多様なデータを客観的かつ詳細に記録できる。

2.2 データ保存機能と検査結果閲覧

被検者の行動データは課題ごとに CSV ファイルとして保存され、検査者は関心に応じてデータを取得し解析することができる。また、検査結果閲覧画面では各課題の正答率とともに課題中の反応時間や刺激提示位置とタップ位置等が表示され、数値やグラフで被検者のパフォーマンスを確認することができる。

3. 自動採点機能の実装

3.1 自動採点機能とデータ保存

新たに実装したシステムでは、課題ごとに保存した CSV ファイルをアプリ内で読み込み、各課題の正答率と反応時間、エラー分類によりスコアを自動計算している。また、自動採点したスコアは CSV ファイルとして出力されアプリ内のデータ閲覧画面に読み込まれる他、検査者が関心に応じて取得できる。

3.2 検査結果フィードバック機能

新たに自動採点システムを実装した目的は、検査者の作業量を軽減し、検査結果処理にかかる時間を短縮することであった。そこで、より効率的に検査結果をフィードバックするため、保存した CSV ファイルか

らスコアと被検者の基本属性を参照しアプリ上で確認する機能を実装した（図1）。

図1 検査結果フィードバック画面

新たなフィードバック画面(図1)では、被検者の基本属性と検査日時、所要時間、スコアの他に、基本属性とスコアに応じた所見を表示している。

4. おわりに

本研究では、認知機能評価・トレーニングを目的とした Android アプリケーション UX-TMT の改良を行っている。本稿では、臨床現場における検査者の作業量を軽減し結果処理にかかる時間を短縮するため、新たに実装したスコア自動採点機能とデータ保存機能、検査結果フィードバック機能について報告した。今後は、実臨床で検査と結果処理を行なっている専門家の意見を参考に本システムの実用性やスクリーニング性能を検討し、さらなる改良を行っていく。

謝辞

本研究を進めるにあたり、ご協力いただいた皆様に感謝の意を表します。

本研究は JSPS 科研費 22K03164 の助成を受けたものです。

参考文献

- (1) 厚生労働省, “認知症施策推進総合戦略(新オレンジプラン)~認知症高齢者等にやさしい地域づくりに向けて~, https://www.mhlw.go.jp/file/06-Seisakujouhou-12300000-Roukenkyoku/nop1-2_3.pdf, (2023年5月9日確認)

- (2) 厚生労働省, “認知症施策推進大綱”, (2019)
- (3) Kokubo Naomi, Yokoi Yuma, Saitoh Yuji, et al.: “A new device-aided cognitive function test, User eXperience-Trail Making Test (UX-TMT), sensitively detects neuropsychological performance in patients with dementia and Parkinson's disease.”, BMC Psychiatry, 18(1), 220 (2018)
- (4) Sirilertmekasakul Chananchida, Rattanawong Wanakorn, Gongvatana Assawin, et al.: “The current state of artificial intelligence-augmented digitized neurocognitive screening test.”, Front. Hum. Neurosci, 17, 1133632. (2023)

没入型 VR を利用した大量調理シミュレータの開発

－自己観察機能(ミラーリング機能)を利用した練習システムの検討－

中山 洋^{*1} 堀端 薫^{*2}

^{*1} 東京電機大学 理工学部 情報システムデザイン学系

^{*2} 女子栄養大学 栄養学部 栄養実践学科

Development of mass cooking simulator using immersive VR

Hiroshi NAKAYAMA^{*1}, Kaoru HORIBATA^{*2}

Tokyo Denki University ^{*1}, Department of Information System design^{*1},

Kagawa Nutrition University ^{*2}, Department of Applied Nutrition ^{*2},

The purpose of this research is to develop a large-scale cooking simulator that enables the acquisition of sufficiently basic skills. In preliminary research, Itou et al. (1) developed a mass cooking simulator using VR. This simulator allows a beginner to train how to move the expert's spatula by following the trajectory of the expert's spatula. As a result of measuring the movement, it was suggested that the learning effect is high. Specifically, we found that novices were closer to the trajectory of experts than previous systems when using this system. However, beginners were not able to see their own body movements objectively. Therefore, a self-observation function (mirroring function) was incorporated into this simulator. As a result, it was suggested that the learning effect of this simulator is improved by being able to practice not only how to move the spatula but also how to move the whole body.

Keyword: 大量調理、シミュレータ、バーチャルリアリティ (VR)、

1. 背景

現在、栄養士資格を取得するには、栄養士資格必修項目をすべて履修し、卒業することが定められている。しかし、給食管理実習などの集団調理実習では、作業を分担して大量調理を行うため、少ない実習実施回数の中で、1人の学生が関わることでできる工程が少なく、実習の全体像を把握できないという問題がある。また、大量の具材を扱うため多くの経費が掛かる。さらに、重労働となる実習の特質上、現状での女子大学や短大における栄養士養成施設での集団調理実習では、女性の実習環境として適切であると言えない。それらの問題を解決するために、Virtual Reality(以下 VR)を用い、大量調理シミュレータを開発した。

前提研究で、伊藤⁽¹⁾は VR を使用した大量調理シミュレータを使用して、熟練者のスパテラの軌跡を用いて、初心者が熟練者のスパテラの動かし方を学習するシステムを作成し、シミュレータの学習効果の向上を目的とする実験を行った。動きを測定した結果として、このシステムを用いたときに、以前のシステムよりも熟練者の軌道に近づくことが分かり、学習効果が高いことが示唆された。しかし、自らの体の動きを客観的に見るができていなか

った。このことからスパテラの動かし方だけでなく、体全体の動かし方を練習できるシステムを開発する必要があると考えられる。

2. 目的

前提研究の課題と前述の背景を受け、VR とモーションキャプチャと自己観察機能を用いて、適切な体の動かし方を学習する練習システムを作成し、シミュレータの学習効果の向上を図ることを目的とした。

3. システム概要

本研究では、シミュレータを使用する際の臨場感を高めるためにVRを用いた。なお、本シミュレータではVRを実現するための装置として、Head Mounted Display(以下 HMD)である、VIVE Pro (HTC社製)を採用した。

なお、本研究のシステム全体像を図1に示す。被験者はHMDを被り、トラッカーのついたスパテラを持って実験を行う。

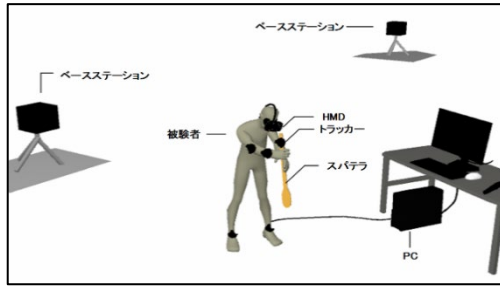


図1 システム全体図

本シミュレータは、チュートリアル機能、練習機能、データ測定機能の3つの機能で構成されている。チュートリアル機能ではVR空間でスパテラを動かす際の手順やポイントを動画、3Dモデル、文字から学ぶことが出来る。練習機能では2つのシーンがある。まず、VR上に自分の動きが映し出される鏡で自分の動きを確認しながら、練習を行える鏡のシーン(図2:左)。次に、鏡のシーンでスパテラを動かしている際に録画されたモーションデータを割り当てた3Dモデルと、熟練者のモーションデータを割り当てた3Dモデルの2つのモデルを比較・観察し体の動かし方の練習を行うことが出来るシーン(図2:右)である。データ測定機能では、回転釜の中の具材を1分間攪拌し、その際のスパテラの座標データ、具材の各面が回転釜と触れている時間で変化する値(以下 接触値)を取得できるシステムとなっている。



図2 練習機能(鏡のシーン、録画の比較・観察シーン)

4. 実験概要

初心者の大学生20名を対象とする。本実験の初心者とは、大量調理を行ったことのない者、かつ本シミュレータを1度も体験したことのない者とした。チュートリアル機能後に、VR上での体の動きを利用した練習機能を持つシステム(以下 新システム)と、チュートリアル後に熟練者の3Dモデルと動画を見て、具材を混ぜることができる練習機能を持つシステム(以下 旧システム)を用いた実験を行う。実験は図3の手順で実施した。

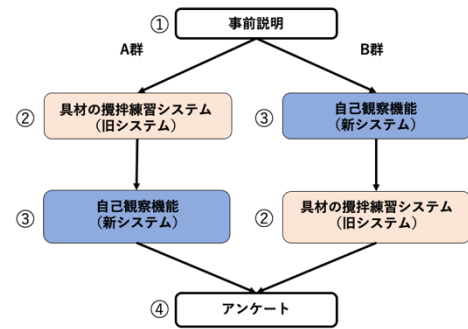


図3 実験フロー

5. 結果

実験で行ったアンケート結果に T 検定を用いて分析を行った。そのアンケート結果を図4に示す。なお、図4の設問内容は「設問1 適切な体の動かし方が身についたと思う」「設問2 楽しく学習することができた」「設問3 適切な体の動かし方が理解できた」「設問5 適切な体の動かし方を習得するのに効果があると思う」「設問6 実際の調理でも適切に体を動かせると思う」「設問7 このシミュレータを利用すればするほど体の動きが上達すると思う」である。

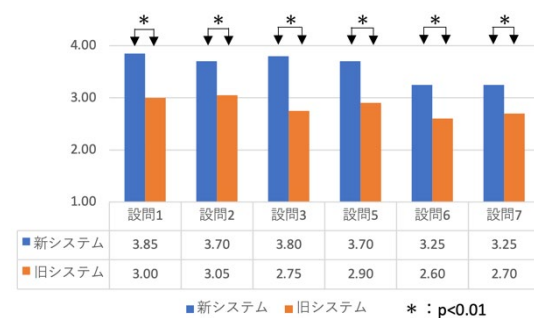


図4 実験結果(アンケート)

6. 考察と今後の課題

アンケートの有意差が見られた回答結果は、全て旧システムより新システムが高評価であった。これは、新システムの自己観察機能を用いた際に、自らの動きを客観的に評価できるため体の動かし方を学ぶことに適していたのだと考えられる。課題として、練習システムの学習効果をより高めるために、フィードバック機能による学習評価機能の追加を検討している。

参考文献

- (1) 伊藤 雅弘、中山 洋:没入型 VR を利用した大量調理シミュレータの開発 -熟練者のスパテラの軌跡を利用した練習システムの開発- ny2020-06 東京電機大学 2020年度卒業論文

学習履歴と教員ポリシー双方を考慮した CBT の難易度分類アルゴリズムの構築

荒関虹希^{*1}, 上野春毅^{*2}, 小松川浩^{*1}

^{*1} 公立千歳科学技術大学大学院 理工学研究科 ^{*2} 公立千歳科学技術大学 理工学部

Construction of a CBT Difficulty Classification Algorithm Considering Both Learning History and Teacher Policy

Koki Araseki ^{*1}, Haruki Ueno^{*2}, Hiroshi Komatsugawa^{*1}

^{*1} Graduate School of Science and Engineering, Chitose Institute of Science And Technology

^{*2} Faculty of Science and Technology, Chitose Institute of Science and Technology

We construct and evaluate a difficulty classification algorithm for a wide variety of exercises and test materials consisting of images and texts managed by each material provider. In the previous study, classification was performed by using test theory to supplement missing values based on learning history. In the present study, we use machine learning, including learner's test results with missing values and the learning policy designated by teachers.

キーワード: 個別最適化, テスト理論, 機械学習, 難易度分類, 教員ポリシー

1. はじめに

近年, 初等中等教育機関での Giga スクール構想を通じて, 学校と家庭を繋ぐ中での個別最適な学習を図ることが求められている. そのための学習支援システムとして, オンライン上で演習やテスト問題を管理し, 個々の学習者に適した教材提供する Computer-based Testing (CBT) の活用が期待されている. その一例として, 文部科学省が提供する CBT システム (MEXCBT) や, 商用の AI ドリルが挙げられる. しかし, MEXCBT⁽¹⁾は出版社や専門家が提供する複数の教材提供が提供した体系的な無償の問題群を, 管理者が一元的に管理して提供するものであり, 適応的に問題を出題する Computer-Adaptive-Testing (CAT) 形式にはなっていないため, 個別最適な学習支援の観点で課題が残る. 一方, 商用の AI ドリルには, CAT 機能は実装されているが, 提供教材がベンダー固有の有償教材となっている. そのため, 各地域特性や学校の置かれている状況が異なる現在の日本の教育事情を考慮すると, 必ずしも現場教員の日頃の教育活動に即した個別最適な学習支援を図れているとはいえない.

現場教育に即した個別最適な学習支援の観点では, 現場に近い多くの教材提供者 (都道府県・市町村単位の教育委員会や複数の学校コミュニティ) が柔軟に無償の CAT 向けの教材を提供できることが望ましい. しかし, 教材を更新する際に, 都度各演習やテストの難易度の設定が必要となり, 現場教員の負担という観点で課題が残る. さらに, 信頼性・妥当性の高い真正な評価を意識したテスト理論に基づく (公開性が担保されている) 難易度の設定を各教育現場レベルで行うことは, 現実的ではない.

そこで本研究では, テスト理論と機械学習アルゴリズムを活用して, 教育現場が日頃から保有する多種多様なテストや演習問題群に対する難易度自動分類アルゴリズムの構築を目的とする. それにより, 教育委員会や学校といった各教材提供者内で運用できる個別最適な学習向けの CAT システムの実現を目指していく.

実証的な研究を進めるため, 現場教員が日頃活用している実際の CAT における問題の難易度の自動調整を想定し, 北海道内に 9 万人規模のユーザを持つ CIST-Solomon⁽²⁾の CAT を基盤システムとして扱う. 当該システムにおけるテストは, テスト理論による CAT のロジック

で稼働している。

2. 先行研究と本研究の位置づけ

2.1 テスト理論

難易度分類に関する先行研究では、テスト理論が採用されることが多い。テスト理論とは、テストに関する知識体系や能力判定のための統計モデルの理論であり、テストの標準化・尺度化を主な目的としている。テスト理論は、項目の難易度・識別力を測定する項目反応理論 (Item Response Theory :IRT) ⁽³⁾や潜在特性を順序尺度にした潜在ランク理論 (Latent Rank Theory :LRT) ⁽⁴⁾が代表的である。

2.2 先行研究の例

金西らの研究⁽⁵⁾では、初年次教育の理系基礎科目を対象とした大規模な演習教材群に対し、CAT を用いた個別最適な学習システムの仕様検討を行った。その際、学習システムにおける問題の作成と解答履歴の収集を行い、IRT を用いた問題の難易度分類を試みた。その結果、IRT を用いて正確な難易度分類を行うためには、様々な理解状況の学習者から学習履歴を収集する必要があるとした。

阿部の研究では、CIST-Solomon の回答履歴と Exametrika⁽⁶⁾ (LRT を実装可能なシステム) を用いて推定された IRP をもとに難易度更新を行った。なお難易度はレベル 1~7 の 7 段階としていた。具体的な阿部のアルゴリズムを図 1 に示す。最初に CIST-Solomon のデータベースからある単元の回答履歴を抽出し、Exametrika を用いて RMP (受験者のランク確率) を推定する。このとき、抽出した回答履歴には欠損値があるため、阿部が作成したシステムでは RMP をもとに正答する確率を仮定し、重み付けを行ったうえで欠損値を補完していた。その後、補完済みの回答履歴 (Excel ファイル) を再度 Exametrika で分析し、推定された IRP を各問題の更新された難易度としていた。

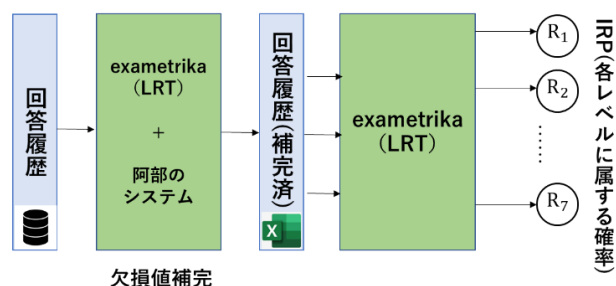


図 1 阿部のアルゴリズム

2.3 本研究の位置づけ

テスト理論では、真正な評価を意識すると各演習問題に大規模なテスト履歴に基づき複数のパラメータ推定が必要となる。一方、近年初等教育での Giga スクール推進や高等教育でのポストコロナでのオンライン教育の普及に伴い、教育機関独自に e ラーニングの教材運用する事例が増えている。こうした学習コミュニティでは、現場教員が適宜教材管理することから、演習問題の学習履歴も均質とならない。そこで、本研究では教材提供コミュニティ単位での教育事情 (内容) に応じた個別最適な学習の提供を可能にするために、テスト理論に基づく CAT に対し、教員の教育ポリシーを反映した機械学習アルゴリズムで置き換えたロジック構築を図り、その有用性を評価する。

本研究では、先行研究が持つ課題点を解決するために、汎用的な実稼働システムである CIST-Solomon の学習履歴を事例として扱い、ループリックを入力パラメータとして追加した全結合型ニューラルネットワーク (Neural Network :NN) のアルゴリズム構築を目指す。このとき、LRT が IRT よりも分類において優れた性能を表すため、阿部の先行研究をもとにアルゴリズムの構築を図る。

3. アルゴリズムの構築と評価

3.1 難易度の定義

本研究では、CIST-Solomon における理解度テスト (CAT) の学習履歴をもとに、難易度分類モデルの作成を行う。そのため、難易度更新の対象として、本学の講義内で理解度テストを例年実施している「アルゴリズムとプログラミング」から、「再帰処理」と「ソートアルゴリズム」の 2 単元を選択した。ただし、この 2 単元は先行研究である阿部のアルゴリズムによって難易度が更新された単元である。そのため、この 2 単元の難易度は、問題を登録した専門家の意見や LRT による推定等、多角的観点から定められたものである。また、2 単元を対象とした理由は、検証の際に単元による相違点を比較するためである。

3.2 難易度分類アルゴリズムの構築

阿部の先行研究では、CIST-Solomon における理解度テストの問題が真正な評価に基づくとは仮定し、LRT を活用することで潜在ランク R1~R7 を求め、最大となる潜在ラン

クを問題の難易度として推定していた。一方で、学生の自習で活用されるシステムでは、学習者毎に回答する問題も異なる。また、適宜教員による問題の追加も行われることを考慮すると、未解答の回答履歴（欠損値）も多い。これに対して、阿部の研究では一般的なNNで用いられる最尤推定による、確率に基づいたデータ補完が行われていた。この場合、LRTの構造自体が教師なしNNであるNeural Network SOM (SOM)⁽⁷⁾と同じであるため、上記のデータ補完を含めて、データ補完を含めた全結合型のNNを用いた機械学習モデリングで表現できる可能性がある。一方で、実稼働システムでは、真正なデータが揃わないデータに対しては、表1に示す教員の教育ポリシー（ループリック）に沿って問題の難易度を決めている。そこで、本研究ではNNの入力パラメータとしてループリックを追加し、教師データを教員により更新された難易度とすることで、全結合型NNによる難易度分類アルゴリズムの構築を図ることとした。

表1 レベルとループリックの対応表

レベル	ループリック
1	初級
2	
3	中級
4	
5	
6	上級
7	

第一に、図2に示すような阿部のアルゴリズムにNNを組み合わせたアルゴリズムを構築した（以下、アルゴリズム1とする）。すなわち、アルゴリズム1ではLRTをNNの入力層として明示的に利用した。さらにループリック情報を入力層に加え、3層構造のNNを構築することで、教員の意見が一定程度反映される難易度更新アルゴリズムとした。

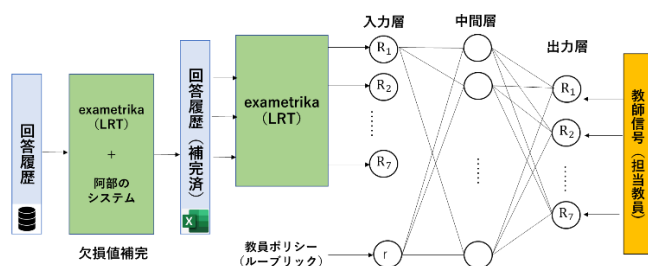


図2 アルゴリズム1

第二に、図3に示すようなExametrikaと阿部のシステムによる欠損値補完とNNを組み合わせたアルゴリズムを構築した（以下、アルゴリズム2とする）。アルゴリズム1では、補完後のデータに対してExametrikaを通してLRTによる潜在ランクを推定しNNの入力に利用した。一方で、アルゴリズム2では、補完後のデータから算出された基本統計量をNNの入力に用いることで、テスト理論に依存しないアルゴリズムを目指した。

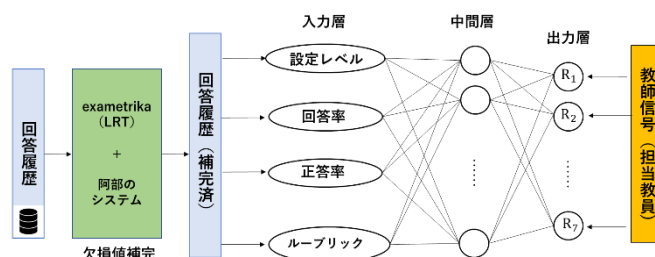


図3 アルゴリズム2

第三に、図4に示すような学習履歴の欠損値補完を行わず、NNのみによるアルゴリズムを構築した（以下、アルゴリズム3とする）。アルゴリズム2では欠損値補完を行ったうえでNNを適用していたが、アルゴリズム3では、Exametrikaによる補完を行わず欠損値が存在するデータから算出された基本統計量をNNに用いた。そのため、LRTに依存しないアルゴリズムとなり、将来的にシステムのみによる自動更新を可能にした。

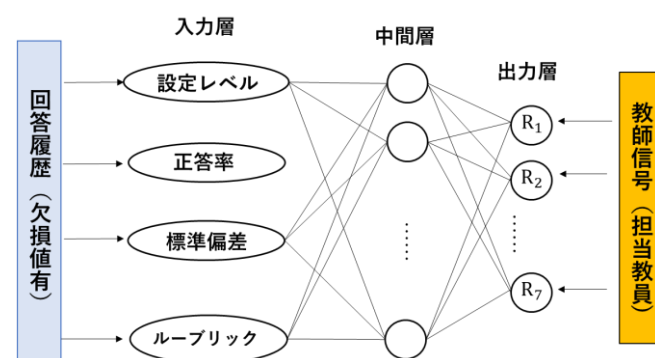


図4 アルゴリズム3

3.3 モデルの作成

前述の阿部のアルゴリズムとアルゴリズム1～3を用いて、「再帰処理」と「ソートアルゴリズム」の2単位において計8種類のモデルを作成した。本研究におけるモデルの教師信号とは、阿部のアルゴリズムによって2021年の解答履歴をもとに分類された問題を、担当教員が単

元ごとに相対評価することで再定義した難易度である。そのため、学習時における入力データは2021年の解答履歴とすることが望ましい。しかし、2021年の解答履歴のみを用いた場合、「再帰処理」の問題数は79件、「ソートアルゴリズム」の問題数は93件のみであり入力データが不足する。そのため、本研究では学習時の入力データを2017～2020年の解答履歴、教師信号を教員により更新された難易度とすることでデータセットを4倍にかさ増した。また、推論時の入力データは2021年の学習履歴とし、教師信号は学習時と同じく教員により更新された難易度とした。この際、2017～2020年の各単元の問題は2021年のものと同質であり、担当教員は主に問題の内容とループリックをもとに難易度を更新しているため、2017～2020年においても担当教員は同様の更新を行うと仮定した。NNのハイパーパラメータに関しては、「再帰処理」のモデル作成時に調整を行い、epoch数を1000、バッチサイズを1として固定した。

3.4 モデルの評価

更新された難易度を評価する際に、観点により難易度の解釈が異なるため、一般的に「正しい難易度」とする指標は存在しない。そこで本研究では、「アルゴリズムとプログラミング」の担当教員に難易度更新における協力を依頼し、その難易度を一時的に「正しい難易度」と定義した。そのため評価における精度とは、開発したモデルにより推定された難易度と担当教員が更新した難易度の一致度を示している。

表2に各モデルによる単元ごとの評価を示す。ここでレベル毎の問題数の分布を測るために、標準偏差も求めた。精度に着目すると、先行研究である阿部のアルゴリズムではどちらの単元も約6～7割の精度であったが、本研究で提案したアルゴリズム1～3のうち、「再帰処理」ではアルゴリズム1・3が約8割、「ソートアルゴリズム」ではアルゴリズム2・3が約7～8割と高い精度を示した。特に、潜在ランクを一切扱わないアルゴリズム3は、2つの単元において最も高い精度を示していた。

標準偏差に着目すると、「再帰処理」ではアルゴリズム1が最も小さく、「ソートアルゴリズム」ではアルゴリズム3が最も小さかった。しかし、実際に分類された問題を分析すると、問題数が0～2問ほどしかないレベルも存在した。そのため、理解度テストを構成するうえでど

のアルゴリズムも適切な更新・分類が行われたとはいえない。

表2 モデルによる単元ごとの評価

アルゴリズム	再帰処理		ソートアルゴリズム	
	精度(%)	標準偏差	精度(%)	標準偏差
阿部のアルゴリズム	65.8	7.12	62.4	14.1
アルゴリズム1	79.8	6.39	54.8	22.6
アルゴリズム2	58.2	7.56	73.1	16.6
アルゴリズム3	84.8	7.68	78.5	10.2

4. 考察と今後の展望

以上より、本研究における精度の観点からは、アルゴリズム3すなわちNNのみで構成されたアルゴリズムが最も優れていたと推測される。これは潜在ランク理論が大規模かつ欠損のないデータを前提に提唱されているものであり、本学におけるCAT形式の理解度テストではデータ欠損やデータ不足が見られるため、十分に作用しなかったためと考えられる。しかし、本研究では入力データとして、阿部の先行研究によって専門家の意見やLRTなどの多角的観点から難易度分類が行われた問題を使用している。そのため、NNのみのアルゴリズムに関しても、入力データを通してLRTが作用している可能性も捨てきれない。

一方で、各レベルにおける問題数の分布の観点からは、担当教員が更新したものと比較すると、レベル間における分布の偏りがあり、なかには1問も属していないレベルが存在するなど、実稼働システムとしての課題が残る。この課題の要因としてモデリングの手法が挙げられる。改善案としては、2021年の解答履歴と担当教員により与えられた教師データのデータセットを増やすことが挙げられる。そのため、本研究では2単元に対してそれぞれモデルを作成したが、教員による難易度分類を行った対象単元を増やし、単元を越えて各レベルの分布が均等になるようにデータを混ぜて学習することで課題解決が見込まれる。また、別の要因としてループリック、すなわち教員ポリシーの学習効果が弱いことが挙げられる。実際に教員が問題の難易度を設定する際には、回答履歴ではなく各問題文の文脈に着目している。そのため、難易度分類をする際に文脈の類似度も考慮する必要があると考えられる。

今後の展望として、BERTを用いて問題文のベクトル表

現を算出し、NN の入力層に組み入れることで難易度分類にどのような影響があるのかを分析していきたい。

参 考 文 献

- (1) 文部科学省. “文部科学省 CBT システム (MEXCBT: メクビット) について” .
https://www.mext.go.jp/a_menu/shotou/zyouhou/mext_00001.html (2023 年 6 月 10 日確認)
- (2) 公立千歳科学技術大学. “e ラーニングシステム CIST-Solomon ” . 公立千歳科学技術大学.
<http://solomon.mc.chitose.ac.jp/CIST-Shiva/Index> (2023 年 6 月 10 日確認)
- (3) 宇佐美 慧, 荘島 荘二郎, 光永 悠彦, 登藤 直弥. 項目反応理論 (IRT) の考え方と実践——測定の質の高いテストや尺度を作成するための技術——. 日本教育心理学会第 60 回総会発表論文集. 2018, 60, 24-25
- (4) 荘島 宏二郎. ニューラルテスト理論—資格試験のためのテスト標準化理論—. 電子情報通信学会誌. 2009, 92, 1013-1016
- (5) 金西 計英, 石田 基広, 戸川 聡, 高橋 暁子. 初年次教育の理系基礎科目を対象にした適応型システムの検討. 日本教育工学会論文誌. 2021, 45, Suppl, 189-192
- (6) 荘島 宏二郎. ” exametrika” . SHOJIMA Kojiro’ s Website .
<http://shojima.starfree.jp/exmk/index.html>
(2023 年 6 月 12 日確認)
- (7) 徳高 平蔵, 藤村 喜久郎. 自己組織化マップとその応用事例. 日本神経回路学会誌. 2003, Vol. 13, No. 3, 147-157