G-41

認知障がい者リハビリテーションにおける 行動映像理解と認知機能自動評価方式の検討

瀬良 大道†　　小川 巧†　　北川 未貴‡　　松田 展†　　大井 翔†‡　　佐野 睦夫†‡

あらまし 高次脳機能障がい者は日常生活の遂行に困難を伴うため、生活行動に基づく認知リハビリテーションが行われており、先行研究では情報処理によるナビゲーションシステムを検討したが、一部で人手での作業を要した。本研究では、片付け行動に着目した認知リハビリテーション支援の一環として、大規模マルチモーダルモデル（LMM）を用いて振り返りコメントを自動生成する手法を提案する。予備実験として、6 名の大学生による片付け動作を撮影し、取得した映像を LMM に入力して 3 つの評価観点に基づいた振り返りコメントを自動生成し、その出力の妥当性を検証した。

キーワード LMM、認知機能障害、リハビリ、映像解析、動作推定

1. はじめに

高次脳機能障がいとは、事故や病気に起因して脳が損傷することにより、記憶障害・注意障害・遂行機能障害・社会的行動障害等の認知障害が現れ、日常生活および社会生活への適応が困難な状態を指す.

深津[1]によると、高次脳機能障がいの支援においては、2001 年以降制度および各支援機関の整備が進んできた一方で、相談事業所に訪れる人の 3 から 4 分の 1 が未診断であること、7 割の就労支援機関がスタッフ不足により高次脳機能障がい者の支援を行っていないこと等、各機関の連携においては問題が依然として残っている.

そこで、我々は高次脳機能障がい者の日常生活動作に基づく認知リハビリテーションを支援する情報システムの検討を行ってきた．大井ら[2]は高次脳機能障がい者 5 例を対象に、調理行動を撮影・解析し、振り返りの対話を提示することで、調理時における対象者の気づきとリハビリに対する意欲を引き出す手法を検証した．また、佐野ら[3]は 2 例に対し片付けの動作における狭義の遂行機能を定義し、それに基づいて映像を解析し振り返りコメントの生成を行った．ただし、これらは主に振り返りの文章の生成において一部手作業を要した．高次脳機能障がい者への支援事業が人員不足の状況にある現在、これらの処理を自動で行うことが求められる.

本研究では、物の片付け動作に基づく認知リハビリテーションにおいて、大規模マルチモーダルモデル（以下、LMM）を用いて、自動的に振り返りの文章を生成する手法を検討する．このシステムによる出力の妥当性を分析した後、リハビリ対象者の気づきとリハビリへの意欲にどのような影響を与えるのかを分析する.

2. 提案方式

本研究の最終目標は、複数のシステムを統合した片付け行動に着目した全自動認知リハビリテーション支援システムに活用することである．認知リハビリテーションに片付け行動を採用した理由は，片付け行動は障害のレベルにかかわらず日常生活に必須の行動であるためであり、実験協力をお願いしているリハビリセンターとの協議により決定した.

† 大阪工業大学大学院, 情報科学研究科

‡ 大阪工業大学, 情報科学部

支援システムの全体図を図 1 に示す．システムに必要な要素は 4 点あり，はじめが VR による部屋の理想状態の提示、その次が行動と物体の認識、その次が行動の自動評価による振り返りコメント生成，最後が振り返りナビゲーションの自動生成である．理想状態の提示では、レンジセンサや環境カメラから獲得される 3 次元仮想環境において、片付けの理想状態を利用者に提案する．行動および物体認識では，実空間上のリハビリ動作をセンサや環境カメラから獲得し，片付けの動作履歴を集積する．自動評価による振り返りコメント生成では，物体認識および行動認識結果を LMM に入力することにより，遂行機能・注意機能・記憶機能といった認知機能を自動で評価するとともに、それに応じたコメントの生成を行う．振り返りナビゲーションにおいては、自動生成した評価およびコメントをリハビリ対象者が理解しやすい形式で提示する．本研究は、上記の内 3 点目に該当する、振り返りコメントの自動生成に焦点を当てる.

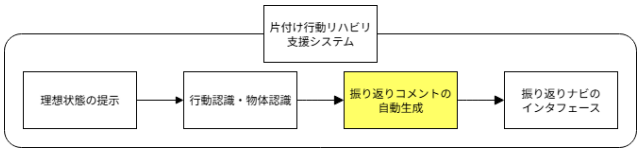

図 1 片付け行動支援システムの全体像

以下に、本研究で提案する手法について説明する.

片付け行動に基づく認知リハビリテーションにおいて、リハビリの動作に対して遂行機能・注意機能・記憶機能に基づいた振り返りコメントを生成するためには、図 1 で示した通り、行動認識および物体認識を用いて対象者の動作を定量化する必要がある．具体的な手法としては、rtmlib[4]を用いて姿勢推定を行い、取得した対象フレームの全身の座標と 1 フレーム前の全身の座標を norfair[5]を用いて比較することで、過去のデータと同一人物かを推定する．これらの処理結果を csv 形式で出力することで、LMM に対し姿勢情報を与えることができる．なお、物体認識手法においては、現在検討を行っている状態である.

座標取得後、LMM を用いて Visual Question Answering を行い、リハビリ対象者の動作を評価し、評価基準ごとの得点とそれを説明した振り返りコメントを生成する．ただしその前に、リハビリ対象者の遂行機能・注意機能・記憶機能に応じた文章を生成する必要があるため、これらの学術的な定義を LMM に与えるための指示文を用意する必要がある．また、片付け動作の評価には評価基準を決定する必

要があるため、どのような動作を対象とするのか、ならびにどのような動作が加点・減点に該当するかを示す必要がある．上記をもとに作成した事前指示文と撮影した動画、姿勢推定で取得した座標を入力として、VQA によって評価基準ごとの得点と振り返りコメントを出力させ、ファイルに保存する．

3. 実験

本研究の目的は、姿勢推定および物体推定によって取得した情報をもとに LMM によって振り返りコメントを生成することである．今回は、予備実験として行動認識および物体認識を使用せず、取得した映像のみを LMM に入力して自動でコメントを生成し、出力の妥当性を確認する．

実験手法について説明する．21～22 歳の大学生 6 名に対して、デスクでの片付け動作の様子を撮影した．机は幅約 1m、奥行き約 0.8m で、協力者ごとに机の大きさは異なる．机には PC のディスプレイとキーボードが共通して配置され、追加でノート PC や書類、飲み物のペットボトル等の片付けるものを配置してもらう．撮影においては、物を適切な位置に整理し配置・収納等の片付けを行ってもらい、開始から終了までの様子を人物が机に向かった状態において、手や物が映るように側面から撮影する．その後、撮影した映像をもとに LMM で振り返りコメントを生成し、Web アプリケーションによって実験協力者に撮影した動画と振り返りコメントを提示し、事後アンケートを実施する．これを各協力者に対して 2 セット実施し、実験を終了した．

撮影には iPhoneXS を使用し、横向きの状態でカメラアプリを使用して撮影した．解像度は 1920×1080、MOV 形式で映像を保存した．撮影した動画は Unsloth が提供している Qwen2.5-VL-7B-Instruct[6]に入力し、振り返りコメントを生成した．動画を入力する際には、必要な内容を満たした文章を生成させるための指示文を付け加えた．指示文は 3 つの条件について得点付けを行う文章を指定した．1 点目は「よく使うものを取り出しやすい場所に片づけたか」、2 点目は「物を決まった場所に片づけたか」、3 点目は「同じものをすべて同じ場所に片づけたか」である．LMM を動作させるために使用した PC の環境は、OS は Windows 11 Home、CPU は 12th Gen Intel(R) Core(TM) i9-12900K、GPU は、NVIDIA GeForce RTX 4090、RAM は 32GB である．Python は 3.11.13、PyTorch は 2.5.1+cu121 を使用し、GPU 処理においては CUDA の 12.1 および cuDNN の 9.1.0 を使用した．

では、実験結果について、実験の様子とそれに対する LMM の出力を 2 件示す。

1 件目として、実験の様子を図 2 に示す．この実験協力者は、不要なものを机下のごみ箱に捨てた後、書籍と電子機器の配置を整えた．なお、個人情報保護の観点から、顔領域にモザイク処理を施している．

図 2 1 例目の実験の様子

続けて、図 2 に対する LMM の出力とその日本語訳は次の通りになった．出力内に示された a.b.c.は、前述の指示文において定義した 3 つの条件に対応しており、a が「よく使うものを取り出しやすい場所に片づけたか」、b が「物を決まった場所に片づけたか」、c が「同じものをすべて同じ場所に片づけたか」である．

['a. 4 points. The person has moved some frequently used items closer to the front, such as the keyboard and mouse, but there is still some clutter around the desk.\n\nb. 3 points. The person has put some items away in specific places, like the box under the desk, but there are still some items that seem to be placed randomly.\n\nc. 2 points. The places where items are put away are not consistently organized; for example, the box under the desk seems to be the primary storage area, but other items are scattered around the workspace.']

a. 4 点 キーボードやマウスなど、よく使うものを手前に移動させたが、机の周りはまだ散らかっている．
b. 3 点．机の下の箱など、特定の場所に片付けてはいるが、まだ無造作に置かれているようなものがある．
c. 2 点．机の下の箱が主な収納場所のようだが、他の物はワークスペースのあちこちに散らばっている．

次に、2 件目として、1 件目と異なる実験例を図 3 と図 4 に示す．この実験協力者は、図 3 では机の上のものを鞄に収納して片付けを完了させたのち、図 4 で机の上に置いて鞄を背負った．これらについても図 2 と同様に、個人情報保護の観点から顔領域にモザイク処理を施している．図 3、図 4 の入力に対し、次の出力が得られた．以下に LMM による出力を抜粋したものとその日本語訳を示す．

図 3 2 例目の実験の前半の様子

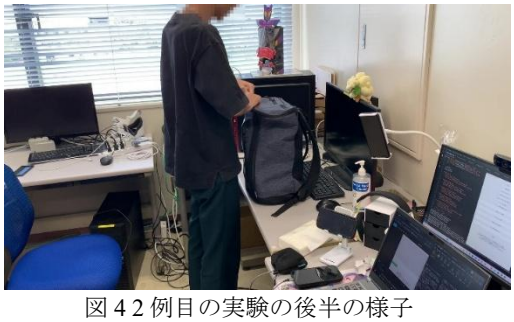

図 4 2 例目の実験の後半の様子

c. 2 points. The places to put things away are not consistent; for example, the backpack is placed on the desk, which is not a designated storage area.

c. 2 点．物の収納場所が統一されていない．例えば、リュックサックが机の上に置かれているが、これは指定された収納場所ではない．

では、図 2 とそれに対する出力コメントについて分析を行う。a においては、but there is still some clutter around the desk とある．キーボードやマウスを整理したことは評価し

ていることを踏まえると、これは図 2 の右下にある紙袋や寝袋の領域を片づけられていないと評価したためと考えられる．ここから、指示文に片付け対象の領域を指定する文章を加える、もしくは、片付け対象の領域のみが映るように撮影環境のセッティングを行う必要がある．b と c より、机下のごみ箱が the box under the desk として、物を収納する箱ととらえていることが読み取れ、不要物を廃棄するという行動に対応していないことが示された．これについては、不要なものをごみ箱に捨てているか、といった評価基準を設けることで対策できると考える．

次に、図 3 および図 4 とそれに対する出力について分析を行う。物を収納した鞄を移動させたことに対し低評価をつけており、収納する場所の位置が変わることに対応していないことがわかる．現行の指示文ではあいまいな要素が多く含まれており、具体的な指示を置き換える必要があると考えられる．

4. 今後の課題

今回は予備実験として、LMM のみで片付け動作の振り返りコメントを自動生成した．その結果、対象外の領域を出力に含めていたことをはじめ、コメントの生成において多数の課題が見つかった．今後、行動認識および物体認識の結果を LMM への入力に追加することと並行して、点数ごとの具体的な目標達成の程度を定義する等、具体的な指標の策定と指示文の見直しを進める必要がある．行動認識においては、rtmlib を用いて YoloX モデルを使用した姿勢推定と、norfair によるキーポイント追跡によって全身のキーポイントの座標を取得し、この結果に基づいて動作の様子を推定する方法を検討している．

また、今回の実験では動画時間は 1 分程度であったが、実際に高次脳機能障がい者の認知リハビリテーションに活用する場合、より長い時間を要すると想定される．LMM に動画を入力する際、長時間の動画を入力に使用すると、処理時間や必要なメモリの量が増大する．この問題に対しては岡野らの研究[7]より、動画から一定の間隔でフレームを取り出して連続性を維持したまま入力することで、限られた計算資源で LMM に行動の様子を与えることが可能であると考える．また、上記の研究は VQA において二者択一の問題を複数回答させることで実空間上の様子を自動で評価した．このように、点数付けの際に、各条件の達成の有無を問う二者択一の質問を複数用意することで、堅牢度の高い出力が得られるのではないかと考えられる．

ほかには、大規模言語モデル（以下、LLM）によるコメント生成も考えられる．行動認識・物体認識により動作を定量化した場合、動画入力がなくともテキストで情報を与え文章を生成することが可能である．今後、取得動画と取得データによる LMM の振り返りテキストと、取得データのみによる LLM の振り返りテキストを比較する、あるいは、同じ LMM でも取得動画を使用せずに生成した振り返りテキストと取得動画を使用して生成した振り返りテキストを比較する等でシステムの構成を検討する余地がある．

また、Özdemir ら[8]は、画像と質問内容に基づいて生成した「質問駆動型キャプション」を LMM で行い、その出力を LLM へのプロンプトとして使用することで、VQA タスクにおける性能を向上させる手法を提案し、その有効性を示した．LMM で動画を分析し、振り返りコメントの生成は LLM で行うといった、LMM と LLM の併用においても、今後の検討課題として挙げられる．

5. まとめ

本研究では、高次脳機能障がい者の片付け行動に着目した認知リハビリテーション支援の一環として、LMM（大規模マルチモーダルモデル）を用いた振り返りコメントの自動生成手法を提案した．従来の研究では、振り返りコメントの作成に一部手作業が必要であったが、人員不足が課題となる支援現場への応用を見据え、本研究ではその自動化を目的とした．

提案手法では、対象者の片付け動作を撮影した映像を LMM に入力し、遂行機能・注意機能・記憶機能の 3 観点から評価点と振り返りコメントを自動生成する．

予備実験では 6 名の大学生による片付け動作を記録し、LMM のみにより生成されたコメントの妥当性を確認した．その結果、対象外の領域まで評価に含めてしまう、曖昧な表現により意図しない評価を行う、片付けの目的やルールに対する誤解など、複数の課題が明らかとなった．

これらの結果を踏まえ、今後は行動認識および物体認識によって動作を定量化し、その情報を LMM に入力することで、より正確で焦点の合ったコメント生成を目指す．また、LMM による出力の堅牢性を高めるために、遂行機能・注意機能・記憶機能の詳細な定義や評価項目の詳細な基準の策定、各評価項目に対し二者択一形式の設問を追加する手法等を検討する必要がある．そのほか、長時間動画に対応するための効率的なフレーム選定手法や、コメント生成タスクにおける大規模言語モデルの利活用も検討する必要がある．

本研究は一部，公益財団法人 JKA の機械振興補助事業の助成（2025M-389）を受けた．

参考文献

[1]深津玲子, “高次脳機能障害者支援の現状と課題―調査研究の結果から―,” 高次脳機能研究（旧 失語症研究）, vol. 44, no. 3, pp. 189-193, 2024.
[2]大井翔, 佐野睦夫, 渋谷美月, 水野翔太, 大出道子, 中山佳代, “高次脳機能障害者の自立に向けた調理行動振り返り支援システムに基づく認知リハビリテーション,” 認知リハビリテーション, vol. 20, no. 1, pp. 51-61, 2015.
[3]佐野睦夫, 中川葵, 小谷凌和, 大井翔, 小山智美, 西野朋子, “高次脳機能障害者に対する掃除行動振り返り支援システムに基づく認知リハビリテーション,” 認知リハビリテーション, vol.22, no. 1, pp. 31-40, 2017.
[4]Tau-J, “rtmlib,” GitHub, 2025-02-28, https://github.com/Tau-J/rtmlib, (参照: 2025-07-23).
[5]tryolabs, “norfair,” Github, 2024-05-01, https://github.com/tryolabs/norfair, (参照: 2025-07-24).
[6]Unsloth, “unsloth/Qwen2.5-VL-7B-Instruct · Hugging Face,” Hugging Face, 2025-04-07, https://huggingface.co/unsloth/Qwen2.5-VL-7B-Instruct, (参照: 2025-07-22).
[7]岡野 将大, 藤井 純一郎, 工藤 雄介, 本田 一彦, “マルチモーダルモデルを用いたアクティビティ調査の自動化手法の提案,” AI・データサイエンス論文集, vol. 5, no. 3, pp. 593-599, 2024.
[8]Övgü Özdemir, Erdem Akagündüz, “Enhancing Visual Question Answering through Question-Driven Image Captions as Prompts,” 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1562-1571, 2024.