

ChatGPTを用いた抽象化と論理展開による複数コメントの段階的なまとめの生成

Stepwise Summarization of Multiple Comments Using Abstraction and Logical Expansion with ChatGPT

伊東 達希^{1*} 砂山 渡² 服部 峻²
Tatsuki Ito¹ Wataru Sunayama² Shun Hattori²

¹ 滋賀県立大学大学院 工学研究科

¹ Graduate School of Engineering, The University of Shiga Prefecture

² 滋賀県立大学 工学部

² School of Engineering, The University of Shiga Prefecture

Abstract: With the development of the Internet, the need to efficiently extract necessary knowledge from vast amounts of information has increased. However, creating summaries that are appropriately abstracted and incorporate new perspectives requires consideration of the information type and intended use. In this study, we propose a stepwise summary generation system that utilizes abstraction and logical expansion of multiple comments with ChatGPT. Through evaluation experiments, we confirmed that the system effectively supports the creation of summaries that suit the intended purpose.

1 はじめに

近年のインターネットの発達により、誰でも簡単に情報の書き込みや閲覧ができるようになった。例えば、Amazonの商品レビューやYouTubeのコメント、オンラインアンケート結果などがある。

しかし、これらの複数のコメントから新たな知識や結論を得るためには、過度に抽象的なまとめや、具体的なまとめのみに着目すると、新しい知識を得ることは難しくなる。また、新しいアイデアが必要な場合には、論理的な展開を含んだまとめが必要とされる。

こういった課題に対し、本研究ではまとめの抽象化と論理展開に着目する。分析の目的やまとめる人によって異なる抽象度、展開度のまとめが必要とされることが考えられるが、両者に着目した意見集約の研究は見当たらない。異なる抽象度と展開度が必要な例として、新商品の開発をするために、商品のレビューをまとめる場合がある。新商品に取り入れる要素を決める場合には抽象度が低い、具体的なまとめが必要とされる。一方、長期的な会社の方針を決める場合には、抽象度が高い、抽象的なまとめが必要とされる。また、それぞれの意外なアイデアを発想するためには、論理展開を多

く含む展開度の高いまとめが望まれる。そのため、抽象度、展開度別のまとめを作成することは新たな知識を獲得するために有効な手段と考えられる。

そこで本研究では、ChatGPTを利用し、複数コメントの抽象化と論理展開による段階的なまとめ生成システムを提案する。複数コメントの自動集約をChatGPTを利用することで行い、段階的に抽象化、論理展開を行うことで、抽象度別のまとめを生成し評価する。段階的なまとめを生成することでユーザーにとって必要な抽象度、展開度を含んだまとめの一覧を作成し、目的に応じたまとめを生成できるシステムの構築を目指す。

2 関連研究

2.1 複数コメントからの要約生成に関する研究

要約生成の研究は、複数コメントの中から重要と考えられるものを抽出する研究と、複数コメントを抽象化した要約を生成する研究がある。

抽出を行う要約生成の研究として、評価視点別レビュー要約のための重要文候補抽出を行う研究[1]がある。この研究はホテルのレビューから観点ごとの重要文のラベルづけを行い、それを学習したSVMを用いて重要

*連絡先：滋賀県立大学大学院工学研究科 電子システム工学専攻
〒522-8533 滋賀県彦根市八坂町 2500
E-mail: on23tito@ec.usp.ac.jp

文の候補を判定し、抽出している。観点も考慮して集約を行うといった手法は本研究と共通しているが、抽出型の要約は、忠実性が高い反面、論理展開による意外な要約を生成することは難しいと考えられる。そのため、本研究では抽出と抽象化のどちらも取り入れた要約を行う。

抽象化を行う要約生成の研究として、教師なしの要約を生成する研究 [2] がある。この研究は、ホテルのレビューの観点ごとの要約を教師なし学習を用いて生成し、Transformer を用いて生成した擬似要約、レビュー、観点の3つの入力から学習し、要約を生成している。要約生成を行う際に観点をを用いる点では共通しているが、本研究では、大規模言語モデルを利用することで、ホテルのレビューに特化せず、より汎用的なまとめ生成システムを構築する。

2.2 大規模言語モデルによる要約生成に関する研究

近年では大規模言語モデルの登場により、要約に特化したモデルでなくても、プロンプトによる指示のみで複数コメントの集約を行うことができる。GPT-3.5 を使って要約生成を行う研究 [3] がある。この研究は、GPT-3.5 に与える入力が長すぎると要約の精度が下がる問題に対し、レビューのデータセットをチャンクごとに区切ることで要約し、要約の真実性を評価する指標の定義を行った。本研究と ChatGPT を用いて要約を行っている点と、評価指標を定義していることが類似している。しかし、ChatGPT のモデルのアップデートにより、ChatGPT が虚偽の内容を回答する頻度が減っているため、本研究では真実性ではなく、まとめの抽象度に着目した評価を導入する。

ニュース記事の要約において、固有表現の追加により、段階的な抽象度の要約を生成する研究 [4] がある。ここでの固有表現とは、人物名や地名、日付や数値といった情報を指す。この研究は、GPT-4 を用いて、ニュース記事内の主要な固有表現から要約を生成し、要約に含まれていない固有表現を順次追加することで、要約の情報量を段階的に増やすシステムを構築した。本研究と段階的なまとめを生成する点では類似している。しかし、この研究では、固有表現の追加を主な手法としており、要約の情報量は変化するが、要約の観点や構造は一定のままである。本研究では、論理的な展開を含めた抽象化を行うことで、異なる観点でのまとめの生成も行う。

3 複数コメントの抽象化と論理展開による段階的なまとめ生成システム

本章では、複数コメントの抽象化と論理展開による段階的なまとめ生成システムについて述べる。

3.1 本研究における抽象化と論理展開

本節では、本研究における抽象化と論理展開について述べる。複数コメントのまとめを生成する目的は、「主要な知識を求める場合」と「新しい知識を求める場合」の二つに分類される。主要な知識を獲得するために「堅実な解釈」が求められる一方、新しい知識を獲得するためには、「論理の飛躍を含めた解釈」が望まれる [5]。そこで本研究では、ユーザの求める適切なまとめを生成するために、「堅実なまとめ」を抽象化によって生成し、「論理の飛躍を含めたまとめ」は論理展開によって生成する。

本研究では、抽象化を「より広い意味に言い換えること」と「結論に必要な部分を省略すること」と定義する。それを実現する処理として次の二つを用いる。一つ目は「目的に対して重要度が高い1箇所（1単語または1文節）を、より広い意味の言葉に言い換える」。二つ目は、「目的に対して重要度が低い1箇所（1単語または1文節）を省略すること」とする。事前に ChatGPT を用いて行った調査においては、一つ目の処理のみで抽象化を行うと、ChatGPT が入力されたコメントを削除せず、明らかに不要な要素も言い換えることで、過度に細部を維持した抽象化を行う傾向が見られた。そこで二つ目の処理を定義することで、不要な要素が適切に除去されやすくなり、段階的な抽象化を行いやすくなった。また、堅実なまとめの生成を促すために、後述するプロンプトにおいて、1度の抽象化処理につき、1つの処理のみを行うことを指定する。

次に、論理展開の定義について述べる。論理展開は、「元のまとめから連想される新しい視点やアプローチを追加すること」と定義する。ここで新しい視点やアプローチは、元のまとめの直接的な言い換えではなく、元のまとめに関連しつつも、新たな前提や異なる観点に基づく内容とする。抽象化は元のまとめから堅実なまとめを生成するのに対し、論理展開は元のまとめから連想される新しい要素を加えることで意外な発想を促す。例えば、「リンゴが落ちた」を順に抽象化を行うと「果物が落ちた」、「ものが落ちた」、「ものが動いた」となるが、論理展開を行うと、「重力が働いた」といった元のまとめの直接的な言い換えではなく、新たな観点を加えたまとめが生成されることを期待する。

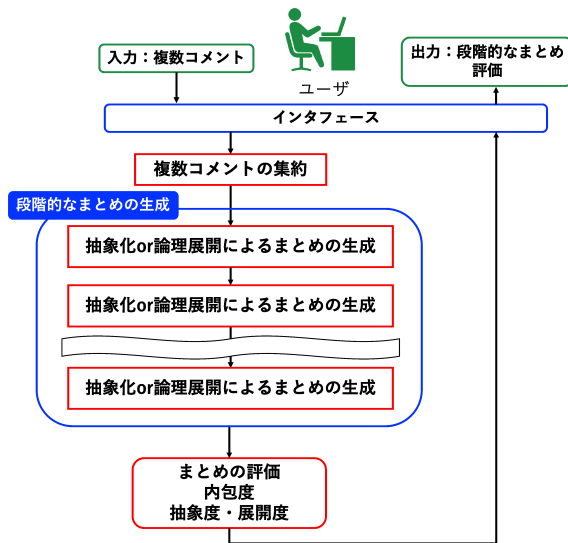


図 1: 複数コメントの抽象化と論理展開による段階的なまとめ生成システムの構成

3.2 システムの構成

図1に複数コメントの抽象化と論理展開による段階的なまとめ生成システムの構成を示す。ユーザは収集した複数のコメントをシステムのインタフェースに入力する。システムは入力された複数コメントを一つに集約し、その集約結果を抽象化、もしくは論理展開を繰り返すことで段階的なまとめを生成する。さらに、それぞれのまとめの評価を行い、インタフェースを通して段階的なまとめと、それらの評価をユーザに提示する。本研究での ChatGPT のモデルは、gpt-4o-2024-08-06 を利用する。

3.3 入力：複数コメント

ユーザが入力する内容は、まとめを行いたい複数コメントと、まとめの目的とする。まとめを行いたいコメントは、1 コメントあたり、20–100 文字程度の意見やレビューとし、コメントの数は2–15 個程度を想定する。これらの想定理由は、複数コメントの集約の際に、文字数やコメント数が多すぎると、重要な要素が抜け落ちてしまう可能性があるためである。本研究では、段階的なまとめ生成に着目しているため、大量のコメントから重要なコメントを抽出する手順はすでに行われているものとする。

また、同じ複数コメントに対しても、観点や目的によって期待する結果が異なるため、まとめの目的を入力する。例えば、「自社商品のレビュー」に対しまとめを作成する場合、「自社商品のレビューから PR 内容を決める」であれば、レビューの中からポジティブな要

表 1: 複数コメントの集約を行うプロンプト（{} 内の文字は変数であり、都度用意された値が代入される）

{目的} という目的において、
{複数コメント} という意見が
指定された文字数になるようにまとめてください。
まとめる方法は、意見の中から共通点を取り出すか、
共通部分を取り出せないもの同士は列挙してください。

入力意見例: りんごを食べたい。トマトを食べたい。
出力意見例: 赤い食べ物を食べたい。

出力は、まとめた意見のみとしてください。
指定する文字数は 100 です。

素に着目する必要がある。一方、「自社商品のレビューから改善策を決定する」であれば、レビューの中からネガティブな要素に着目し、まとめを作成する必要がある。他にも、作成するまとめの方向性の違いや、まとめを行う立場の違いを目的に含むことを想定する。

3.4 複数コメントの集約

本システムでは、入力された複数コメントを ChatGPT を用いて一つに集約する。ここでは、後の処理の段階的な抽象化と論理展開を行うため、可能な限り具体的な情報を保持した集約を行う必要がある。そこで、集約を「意見の中から共通点を取り出すか、共通点を取り出せないもの同士は列挙してください。」と指示することで、入力された複数コメントの要素を可能な限り保持した集約を行う。複数コメントの集約を行うプロンプトを表1に示す。また、具体的な入出力例も含める one-shot プロンプトによる精度の向上も図る。

文字数の指定をしなければ、過度な抽象化が行われ、短い集約結果が返答されることがあったため、複数コメントを集約するのに十分と考えられる文字数として 100 文字を指定する。

3.5 コメント集約結果の抽象化と論理展開による段階的なまとめの生成

本システムでは抽象度と展開度の異なるまとめを生成するために、段階的な抽象化と論理展開を行い、まとめを生成する。また、生成されたまとめが適切あるかを評価するために内包度、抽象度、展開度の3つの指標を提案する。これらの指標を導入することで、ユーザがインタフェース上で目的にあったまとめを選択しやすくなると考える。

表 2: 抽象化を行うプロンプト（{ } 内の文字は変数であり、都度用意された値が代入される）

{ 目的 } という目的において、{ 集約結果・まとめ } という意見を 1 回だけ抽象化してください。
抽象化では、次の A か B のいずれか片方のみを選択して 1 回だけ実行してください。
A) 目的に対して重要度が低い 1 箇所（1 単語または 1 文節）を省略する。
B) 目的に対して重要度が高い 1 箇所（1 単語または 1 文節）を、より広い意味の言葉に言い換える。
抽象化の結果のみを出力してください。
##抽象化例
美味しいりんごが食べたい。 → りんごが食べたい
出力 りんごが食べたい
文字数は 30 文字を目安にしてください。

3.5.1 抽象化と論理展開の繰り返しによるまとめの生成

前節で得られた集約結果に対し、「抽象化」と「論理展開」の 2 つのプロンプトを繰り返し与え、ユーザの目的に合わせたまとめを生成する。繰り返しの回数はユーザが設定できるが、繰り返しの回数が少なければ、十分な抽象化や論理展開が行われず、回数が多ければ ChatGPT の返答に時間がかかる。これまでの複数回の試行により、処理時間と要約の質のバランスがとれた設定として、デフォルトの組み合わせを「抽象化、抽象化、論理展開、抽象化、抽象化」の 5 回とする。

「抽象化」は堅実なまとめを行うために、3.1 節で述べた「言い換え」と「省略」の二つの処理として明示的に記述し、1 度の抽象化で一つの処理のみを行う指示をする。抽象化を行うプロンプトを表 2 に示す。具体例として、「おいしいりんごが食べたい」に対して、「省略」の処理を行うと「りんごが食べたい」となる。抽象化の結果のみを出力させることで、抽象化の理由や行った処理の説明といった、意図しない返答を制限する。

一方、「論理展開」は意外なまとめを行うために、元のまとめから連想される新しい視点やアプローチを追加することを指示する。また、論理展開においても、意図しない余計な説明を避けるために、新しい意見のみを出力する指定を行う。論理展開を行うプロンプトを表 3 に示す。例えば、「日常からの脱却」というまとめを展開すると、「非日常の体験」となり、さらに展開すると「魔法が使える」というまとめが得られる。抽象化とは異なり、論理展開を行うプロンプトを用意することで、新たな視点を加えたまとめが生成されることを期待する。

また、それぞれのプロンプトにおいて、one-shot も

表 3: 論理展開を行うプロンプト（{ 目的 } は変数であり、入力された目的が代入される）

{ 目的 } という目的において、与えられる意見から一度だけ連想的な論理展開を行い、新しい意見の一つ答えてください。
論理展開は、元の意見の要点を維持しつつ、新しい視点やアプローチを取り入れてください。
出力は論理展開してできた新しい意見のみとしてください。
具体例
日常からの脱却 -> 非日常の体験
非日常の体験 -> 魔法が使える
文字数は 30 文字を目安にしてください。

しくは few-shot プロンプトを与え、意図しない過度な抽象化や論理展開を防ぎ、より一貫した出力を促す。さらに、まとめの文字数の指定には、Yahoo! ニュース記事の見出しが最大 15.5 文字であることと、BBC ニュースの日本語記事の要約のデータセット [6] の訓練データの平均文字数が約 96 文字であることを参考にし、両者の中間であり、かつ読みやすさを考慮して 30 文字と設定する。

「抽象化」と「論理展開」の組み合わせの割合を変化させた場合のまとめの内容について述べる。繰り返しの回数を固定した場合には、抽象化の割合が高いほど、元のコメント集合の内容を多く保持したまとめが生成される。一方、論理展開の割合が高いほど新しい視点やアプローチが加わるため、元のコメント集合から離れた内容を持つまとめが生成される。これらの組み合わせを自由に設定することで、目的に応じた多様なまとめの生成を行う。

3.5.2 まとめの内包度の評価

本項では前項で生成されたまとめの内包度の評価について述べる。内包度は、システムに入力されたコメントの個数のうち、まとめに使用されたコメントの個数を評価する指標とする。まとめに含まれているかの判断基準は状況や目的によって変化するため、ChatGPT にプロンプトを与え、評価を行った。内包度の評価を行うプロンプトを表 4 に示す。ChatGPT がそれぞれの複数コメントにまとめが含まれるかどうかを「0」と「1」で回答する。初期検証において、リストの長さが入力のコメント数と異なることがあった。その問題に対処するため、総コメント数をプロンプトに含めることで、正しいリストの長さの出力を促す。

内包度 *Entailment* は式 (1) によって定義される。ここで、 n をコメント集合のコメントの総数、 E_k を Chat-

表 4: 内包度の評価を行うプロンプト（{ } 内の文字は変数であり、都度用意された値が代入される）

{ 目的 } というまとめの目的において、
{ 入力された複数コメント } という意見のそれぞれが、
{ まとめ } というまとめに
含まれているかどうかを判断してください。
少しでも含まれている場合は「1」、
含まれていない場合は「0」として、
意見数に合わせて回答してください。
回答には「1」または「0」のみを使用し、
リスト形式で回答してください。
1 行 1 意見とし、意見数は { 総コメント数 } です。
意見数が 6 の時の出力例
[0,1,0,1,1,0]

GPT による返答のリストの各要素とする。

$$Entailment = \frac{\sum_{k=0}^n E_k}{n} \quad (1)$$

内包度の値は $0 \leq Entailment \leq 1$ の範囲を取り、1に近いほど、まとめが入力された多くのコメントを反映していること意味する。一方、0に近い場合は、まとめが複数コメントの内容をほぼ含んでいないことを示す。

ただし、内包度が高ければまとめの質が良いとは限らない。例えば、抽象化を行わず、入力コメントをそのまま羅列したものに対する内包度は1に近づくが、まとめとして適切とはいえずらい。そのため、内包度はまとめの反映率を測る指標として利用するが、他の指標との組み合わせも必要である。

3.5.3 まとめの抽象度と展開度の評価

本研究において、3.1節で抽象化を「より広い意味に言い換えること」と「結論に必要な部分を省略すること」と定義した。これにより、抽象化後のまとめの意味集合は、抽象化前のまとめの意味集合と比較すると拡大すると考えられる。この拡大率を定量的に評価するために、コサイン類似度を用いる。

また、論理展開は、元のまとめの一部を削減しつつ、新たな要素を追加することを含む。この際、論理展開による意味集合の変化は、削減された要素と追加された要素の差分として捉えられる。そのため、論理展開後のまとめと元のまとめとの共通部分の大きさは、類似度として評価できると考えられる。

それらを踏まえて本研究では、抽象度と展開度の定量的評価のために、Embeddingを用いて各まとめをベクトル化し、コサイン類似度を算出する。この手法により、抽象化による意味の拡張度合いや、論理展開による要素の変化を数値的に測定することが可能となる。

Embedding とは、単語や文章を意味ベクトルに変換する手法である。これを行うことで、ベクトル同士のコサイン類似度を求めることができ、単語や文章の類似度を計算できる。本研究では、未知語への対応や、単語ではなく文章をベクトルに変換できること、APIを用いて簡単に利用できるため、OpenAI の Embedding[7] を利用した。このモデルでは、Embeddingを行うことで、文章が 1,536 次元のベクトルに変換される。本システムでは、2025 年 1 月時点で最新モデルの text-embedding-3-large を使用する。

前項で述べた通り、本システムでは、まとめの生成手法を、抽象化と論理展開の2つから選択し、それを繰り返すことで段階的なまとめを生成する。抽象度と展開度の評価においては、まとめ生成手法で抽象化が選択された場合のコサイン類似度が抽象度に積算され、論理展開が選択された場合のコサイン類似度が展開度に積算される。

また、コサイン類似度が高い場合、抽象化による意味集合の拡張度合いや論理展開の変化の度合いは低い。一方で、コサイン類似度が低い場合、直前のまとめからの変化が大きく、抽象化や論理展開の度合いは高い。したがって、コサイン類似度の積を1から差し引くことで、直感的に理解しやすい指標を設計する。

具体的には、各まとめ生成 n において、「抽象化」が行われた場合の抽象化前後のまとめのコサイン類似度を a_n とし、「論理展開」が行われた場合の論理展開前後のまとめのコサイン類似度を b_n とする。ただし、論理展開が選択された場合、そのステップでの a_n の値が未定義となるため、変化なしを意味する $a_n = 1.0$ を代入する。同様に、抽象化が選択された場合の b_n の値も未定義となるため、 $b_n = 1.0$ を代入する。この定義に基づき、全体の抽象度 $Abstractness_N$ を以下の式 (2) で定義する。

$$Abstractness_N = 1 - \prod_{n=1}^N a_n \quad (2)$$

同様に、展開度 $Expansion_N$ も式 (3) を用いて求める。

$$Expansion_N = 1 - \prod_{n=1}^N b_n \quad (3)$$

ただし、複数コメントを集約した結果における a_0 と b_0 は、コサイン類似度を計算する基準が存在しないため、それぞれ 1.0 を代入する。これにより、初期状態では $Abstractness_0 = Expansion_0 = 0.0$ となる。

図 2: 複数コメントの段階的なまとめ生成インタフェース

3.6 複数コメントの段階的なまとめ生成インタフェース

複数コメントの段階的なまとめ生成インタフェースを図 2 に示す。

画面上部でまとめ生成方法を選択し、中央部で入力を行う。また、画面下部で出力を確認することができる。出力は、一連の処理で生成されたまとめと指標を表示する。指標を全て数字で提示すると視認性が低くなるため、内包度は、4 段階の星の数でユーザに提示する。星の数が多いほど、内包度が高いことを表す。また、集約結果との類似度を視覚的に示すため、まとめのフォントの色の濃淡で提示し、抽象度、展開度はテキストで表示する。

4 複数コメントの抽象化と論理展開による段階的なまとめ生成システムの評価実験

本章では、複数コメントの抽象化と論理展開による段階的なまとめ生成システムが、目的に応じたまとめを得るために有効かを検証した実験について述べる。

4.1 実験方法

本節では、実験方法について述べる。実験で設定した、2つのコメント集合を集約する際の各目的を表 5 に示す。2つのコメント集合と、それらをまとめる 3つの目的に対して、16名の理系大学生、大学院生の被験者に、目的に応じたコメント集合のまとめを、合計 6 回作成してもらう実験を行った。1回の作成の手順は、以下の通りとした。

表 5: 実験で設定した、2つのコメント集合を集約する際の各目的

	動画コメント	商品レビュー
目的 1	動画の特徴を友人に一言で簡潔に伝える際の説明	タオルの特徴を友人に一言で簡潔に伝える際の説明
目的 2	動画を投稿したチャンネル主として、この動画の良さを一言で簡潔に伝えるための説明	タオルの商品開発の視点で、このタオルの良さを一言で簡潔に伝えるための説明
目的 3	動画の特徴を生かした新たな方向性の動画を作る際に、アイデア生成の元になる簡潔なまとめ	タオルの特徴を生かした新たな方向性の日用品を作る際に、アイデア生成の元になる簡潔なまとめ

1. コメント集合をまとめる目的を提示した上で、コメント集合を読んでもらい、そのまとめを作成する。
2. 提案システムによって作成されたコメント集合のまとめを提示し、適切なコメントを選択してもらう。
3. 1. と 2. の内容を踏まえて、再度、コメント集合のまとめを作成してもらう。

実験で用いた目的 1 および目的 2 はサービスを受ける側とサービスを提供する側の観点に分けられる。また、堅実なまとめを必要とする目的 1 と 2 に対し、意外なまとめを必要とする目的 3 を設定した。

コメント集合として用いた動画コメント、および商品レビューは YouTube での再生回数が多い動画、Amazon の売れ筋ランキングで上位の商品を選び、それぞれのコメント集合は評価順で上位 10 件程度抽出した。

手順 2 において、本実験では、抽象度と展開度の指標がユーザの役に立つかを評価するため、内包度の表示は行わず、抽象度と展開度のみを被験者に提示した。また、システムによるまとめ生成の組み合わせをあらかじめこちらで用意することで、システムの利用と同等の結果が得られると考え、まとめの提示、回答は Google forms 上で行った。実験で用意したシステムのまとめ生成手法の組み合わせを表 6 に示す。本システムでは、段階的なまとめを複数回の ChatGPT の応答から得ているが、1 度の応答で段階的なまとめを得ることができれば、システムの待ち時間およびコストを削減できる。そのため、1 回の返答で 1 つの集約と 5 段階の抽象化されたまとめを出力する指示をした ChatGPT による出力も用意した。1 度のリクエストで段階的なまとめを行うプロンプトを表 7 に示す。それぞれの組み合わせにおける 6 個のまとめを 1 セットとして、その

表 6: 実験で用意したシステムのまとめ生成手法の組み合わせ (0 は意見集約, 1 は抽象化, 2 は論理展開, bulk GPT は 1 度の返答で 6 個のまとめを返答する ChatGPT を表す)

組み合わせ
[0, 1, 1, 1, 1, 1]
[0, 2, 1, 1, 1, 1]
[0, 1, 1, 2, 1, 1]
[0, 1, 2, 1, 2, 1]
[bulk GPT]

表 7: 1 度のリクエストで段階的なまとめを行うプロンプト ({ 目的 } は変数であり, 入力された目的が代入される)

{ 目的 } という目的において、与えられる意見を一つの意見に集約し、その集約した意見を少しずつ抽象化することで、抽象度の異なるまとめ意見を 5 個生成してください。集約された意見、抽象化意見 1、抽象化意見 2、抽象化意見 3、抽象化意見 4、抽象化意見 5 の計 6 個の意見を生成してください。

セットをランダムな順番に並べたフォームを作成した。

4.2 実験結果と考察

4.2.1 選択されたまとめの抽象度と展開度からの考察

本項では、選択されたまとめの抽象度と展開度について考察する。手順 2 で選択されたまとめの展開度 (被験者平均) を図 3 に示す。図 3 から、目的「1. 友人に特徴を伝える」において、どちらのコメント集合であっても、抽象度の被験者平均が約 0.08 であった。このことから、「友人に特徴を伝える」という目的に対しては、その特徴をある程度具体的に伝える必要があったと考えられる。

また、目的「2. 制作者として良さを伝える」の商品レビューにおいては、抽象度の被験者平均が 0.18 であった。これはある程度の抽象化を行うことで、商品の良さを漏れなく伝えるためだと考えられる。動画コメントにおいては、被験者平均の値が 0.08 と小さかった。これは、動画の登場人物である、「じいじ」と「ばあば」を用いた抽象化の行われていないまとめが多く選択されることで、抽象度が下がったためであると考えられる。このことから、「制作者として良さを伝える」という目的に対しても具体的なまとめが必要であったと考えられる。

目的「3. 新しいアイデアを生成する」においては、どちらのコメント集合であっても、抽象度の被験者平均が 0.25 以上であった。このことから、「新しいアイデア

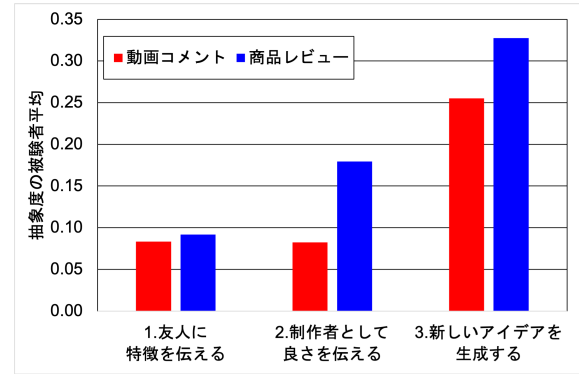


図 3: 手順 2 で選択されたまとめの抽象度 (被験者平均)

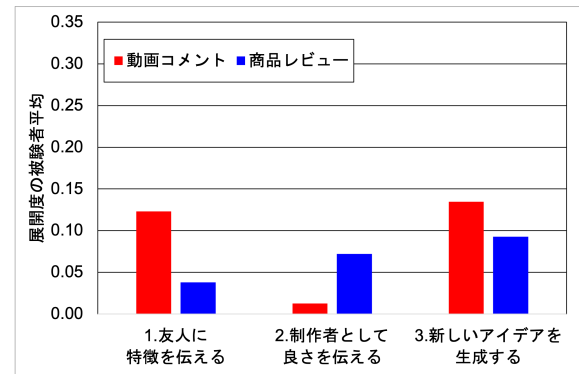


図 4: 手順 2 で選択されたまとめの展開度 (被験者平均)

の生成」に際しては、情報を抽象化する必要性が他の目的と比較して高かったと考えられる。

以上のことから、堅実なまとめを必要とする目的 1 と 2 に対し、意外なまとめを必要とする目的 3 とでは要求される抽象度が異なることを確認した。よってコメント集合や目的が異なれば、異なる抽象度のまとめが必要となる可能性があり、様々な抽象度のまとめを提示することには意味があると考えられる。

同様に、手順 2 で選択されたまとめの展開度 (被験者平均) を図 4 に示す。図 4 より、どの目的においても展開度の被験者平均はおおよそ 0.1 前後であり、顕著な差は見られなかった。このことから、本実験においては、被験者は論理展開を加えたまとめよりも、堅実なまとめを選ぶ傾向があったと考えられる。今回の評価実験では「一言で簡潔に伝えるための説明」といった指示であり、具体的な文字数の指定がなかったことに加え、複数選択可であった。それにより、多くの文字数でより多くの情報を持つまとめ (低抽象度, 低展開度) が選択される傾向が強まったと考えられる。

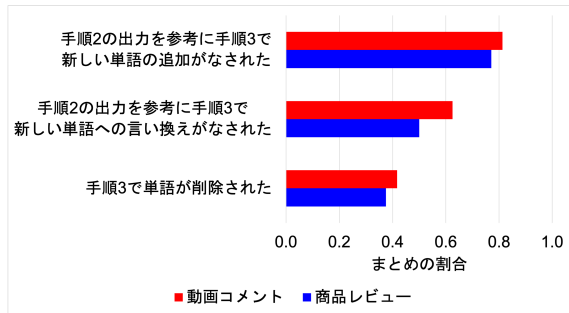


図 5: 手順 1 から手順 3 におけるまとめの変化の重複あり分類の結果

4.2.2 まとめの提示により作成されたまとめの変化についての考察

この項では、まとめの提示により作成されたまとめの変化について考察する。前節の手順 1 と手順 3 において、被験者ごとのまとめの変化を「新しい内容の単語を追加」「内容を維持したまま、新しい単語を用いて言い換え」「単語の削除」の 3 つに重複を許して分類した。そのコメント集合ごとの割合を図 5 に示す。ここから、約 8 割のまとめにおいてシステムによるまとめの内容を用いた単語の追加が行われ、およそ 6 割のまとめが内容を維持したまま、単語の言い換えを行った。このことから、被験者はシステムによるまとめを参考にしつつ、自らの表現に適した単語を追加、修正することができたと考えられる。

また、約 4 割のまとめにおいては、単語の削除が行われた。これは、システムによる抽象化されたまとめを参考に、目的に対して不要な内容を整理できたためであると考えられる。特に、手順 1 でのまとめの文字数が多い被験者においてこういった変化が多く見られた。

これらを踏まえ、システムによる段階的なまとめの提示は、まとめる目的に対して適切な内容を追加し、不必要な内容を削除しながら、ユーザにとって適切な表現へと修正する支援を行える可能性があると考えられる。したがって、本システムはまとめる目的に応じた柔軟なまとめ作成を支援する方法の一つとなりうる。

5 おわりに

目的に合わせた段階的なまとめを生成するために、複数コメントの抽象化と論理展開による段階的なまとめ生成システムを提案した。抽象化と論理展開を定義し、ChatGPT を用いた段階的なまとめ生成によって堅実なまとめと意外なまとめを生成し、それらを評価するシステムを提案した。

また、段階的なまとめの提示がユーザにとって目的にあったまとめの作成に有効であるかを確認する評価

実験を行った。その結果、目的ごとに必要とされる抽象度が異なることがわかった。また、まとめの提示によりユーザにとって必要な要素を追加し、ユーザにとって適切な表現へ言い換える支援として機能する可能性を示せた。

参考文献

- [1] 小池惇爾, 松吉俊, 福本文代: 評価視点別レビュー要約のための重要文候補抽出, 第 18 回言語処理学会年次大会, pp.1188–1191, (2012)
- [2] Amplayo, Reinald Kim, Angelidis, Stefanos, Lapata, Mirella: Aspect-Controllable Opinion Summarization, Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021), pp.6578–6593, (2021)
- [3] Bhaskar, Adithya, Alex Fabbri, and Greg Durrett: Prompted opinion summarization with GPT-3.5, Findings of the Association for Computational Linguistics (ACL 2023), pp.9282–9300, (2023)
- [4] Griffin Adams, Alex Fabbri, Faisal Ladhak, Eric Lehman, and Noémie Elhadad: From Sparse to Dense: GPT-4 Summarization with Chain of Density Prompting, In Proceedings of the 4th New Frontiers in Summarization Workshop, pp.68–74, (2023)
- [5] 砂山渡: フリーソフト TETDM で学ぶ実践データ分析, コロナ社, (2020)
- [6] Mandy Guo, Zihang Dai, Denny Vrandečić, and Rami Al-Rfou: Wiki-40B: Multilingual Language Model Dataset, In Proceedings of the Twelfth Language Resources and Evaluation Conference, pp.2440–2452, (2020)
- [7] OpenAI Platform Vector embeddings: ([URL](https://platform.openai.com/docs/guides/embeddings/)) <https://platform.openai.com/docs/guides/embeddings/> (2025 年 1 月 23 確認)

ChatGPTを用いて比較の観点に応じた特徴を出力する 2つの事柄の比較支援システム

A Support System for Comparison of Two Things that Outputs Features According to The Viewpoint of Comparison using ChatGPT

谷口 拓紀^{1*} 砂山 渡² 服部 峻²
Takuki Taniguchi¹ Wataru Sunayama² Shun Hattori²

¹ 滋賀県立大学大学院 工学研究科

¹ Graduate School of Engineering, The University of Shiga Prefecture

² 滋賀県立大学 工学部

² School of Engineering, The University of Shiga Prefecture

Abstract: 意思決定に向けたデータ分析では、選択肢を比較する分析が必要な場面がある。比較は共通の観点により行う必要があるが、世の中の情報は観点が整理されているとは限らず、比較の目的によって観点が変わることもある。そこで本研究では、2つの事柄の、比較の目的に対応した共通の観点に基づいて差分点と共通点を提示し、比較を支援するシステムを提案する。評価実験の結果、提案システムを用いることで、2つの事柄の違いと共通点をより多く挙げやすくなることを確認した。

1 はじめに

世の中のデータ利活用による意思決定の必要性に伴って、データ分析の需要が増加している。意思決定に向けた分析では、複数の選択肢を候補として比較するための分析が必要になる。例えば、日常生活において家電を購入する際、複数の候補となる商品を比較して検討することがよく行われる。また、複数の商品を比較する際には、一般的に比較の観点が必要であり、各種カタログにおいても、複数の製品のスペックが項目ごとに表でまとめられているケースが多い。

しかし家電など、比較されることが想定されているものについては、あらかじめ比較しやすい形で情報が整理されているが、世の中の任意のものについては、必ずしも比較のための情報が用意されているとは限らない。また、例えば「馬」を動物として「牛」と比較する場合と、「馬」を乗り物として「自動車」と比較する場合とでは、比較の観点が異なってくるため、あらかじめ比較のための情報を用意しておくことが難しいケースも考えられる。

そこで本研究においては、複数の選択肢を比較する場面を想定して、その中で特に2つの事柄の比較を支

援するシステムを提案する。本システムでは、大規模言語モデルの一つである ChatGPT[1] を用いて選択肢についての情報を動的に集めた上で、共通の観点に基づいて、比較する2つの事柄の差分点と共通点を提示する。なお、本研究においては、比較の「観点」を、2つの事柄を比較するための軸となる項目として定義する。また、観点に基づく2つの事柄の違いを「差分点」、一致する点を「共通点」と呼ぶ。

なお、生成AIを利用したシステムの特性上、ハルシネーションの発生を避けることができないため、内容の最終的な真偽は、ユーザ本人が確認する必要がある。

以下本論文では、2章で関連研究について述べ、3章で提案する2つの事柄の比較支援システムの詳細について述べる。4章で提案システムの効果を検証した評価実験について述べ、5章で本論文を締めくくる。

2 関連研究

2.1 大規模言語モデルが持つ知識の利用に関する研究

大規模言語モデルが持つ内部知識を効果的に活用することで応答の正確性を向上させる研究[2]がある。この研究では、モデルが持つ知識を活用することに焦点を当てた手法を提案しており、具体的な分析には用い

*連絡先：滋賀県立大学大学院工学研究科 電子システム工学専攻
谷口 拓紀
〒522-8533 滋賀県彦根市八坂町 2500
E-mail: on23ttaniguchi@ec.usp.ac.jp

ていない。本研究においては、大規模言語モデルが持つ知識を2つの事柄を比較するために用いる。

不足している知識を大規模言語モデルの内部知識で補うことでヒューマンエラーに関する対策の立案に活用する研究 [3] や、モデルの内部知識とモデルとは無関係の外部知識を適切に相互作用させることで精度を向上させる研究 [4] がある。これらの研究においては、モデルが持つ知識に加えて外部の知識を利用することを想定しており、何らかの形で適切な外部知識を用意する必要がある。本研究においては、外部知識を必要とせず、大規模言語モデルが持つ知識のみを利用して2つの事柄の比較を支援するシステムを構築する。

2.2 比較分析を支援する研究

経営工学分野の複数の学会の論文から特徴を表す単語を抽出することで学会の特徴を分析する研究 [5] や、複数の国の新聞サイトの記事から特徴的な単語を抽出することで各国のニュースの比較分析を行う研究 [6] などがある。これらの研究においては、対象のテキストから主に頻出単語を特徴的な単語として抽出することで、その比較を促している。しかし、観点ごとに単語が整理されていないため、単語を見た人がどのような共通点や差分点があるかを、一から判断する必要がある。

観点をを用いた情報の比較を支援する研究として、印象に関する評価軸を用いて複数のニュースサイトの報道傾向の分析を支援する研究 [7]、ネットオークションやオンラインショッピングの商品の特徴を抽出して観点ごとにまとめてユーザに提示する研究 [8, 9] などがある。しかし固定の観点をを用いた場合、想定する比較の目的以外で比較を行うことは難しくなる。そこで本研究においては、比較の目的に応じた観点を動的に用意した上で、その観点に基づく差分点と共通点をユーザに提示するシステムを構築する。

また、大規模言語モデルを用いて、文章の埋め込み表現を求めた上で、文書をクラスタリングして比較分析を行う研究 [10] や、画像分類タスクに対して、画像の要約テキストとそれによるカテゴリ間の比較説明を生成して用いることで分類精度を向上させる研究 [11] がある。しかし、ChatGPT などの生成系 AI によって、比較につながる出力を直接得て、比較を容易にする研究は見られない。本研究においては、ChatGPT が事前に学習している知識を用いることで、任意の2つの事柄の比較を支援するシステムを構築する。

3 2つの事柄の比較支援システム

本章では、ユーザが特定の目的に対して2つの事柄を比較する場面において、同一の観点に対して、2つ

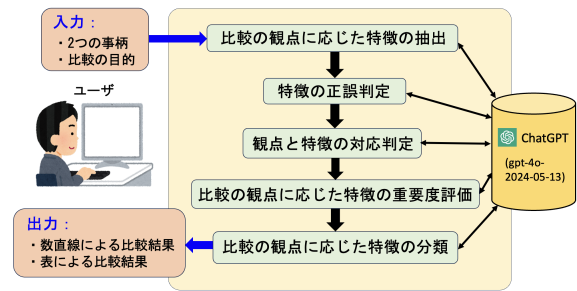


図 1: 2つの事柄の比較支援システムの構成

の事柄それぞれの特徴を対比できる形で提示する比較支援システムについて述べる。

3.1 2つの事柄の比較支援システムの概要

図 1 に2つの事柄の比較支援システムの構成を示す。ユーザはまず、比較したい2つの事柄と、比較の目的を入力する。システムは、入力された目的と事柄をもとに、比較の観点と、2つの事柄の差分点、共通点を抽出する。次に、抽出された差分点と共通点が正しいかどうかを確認した上で、それらが抽出した観点に対応しているかを確認する。その後、抽出された観点が、目的に対してどの程度重要かを判定するとともに、比較が数値的に行える観点を抽出する。最後に、数値的に比較が行える観点は数直線として、それ以外の観点は表形式でまとめて2つの事柄の差分点と共通点を出力する。これらの各処理部分においては、ChatGPT を利用する。なお、本研究では、ChatGPT のモデルとして「gpt-4o-2024-05-13」を用いた。

3.2 入力：2つの事柄と比較の目的

システムには、比較対象となる2つの事柄と比較の目的を入力する。

比較の対象となる2つの事柄は、ChatGPT が事前に学習できている内容で、いわゆるハルシネーション（嘘）が少ないと考えられる事柄とする。言い換えると、ChatGPT が学習に用いたインターネット上の情報に、パターン化されるほどの多くの情報が存在する事柄とする。たとえば、専門性がない一般名詞や、世の中で多く使われている固有名詞などが比較対象となる。

また、同じ2つの事柄を比較する際でも、比較の目的が異なれば比較の観点が異なるため、どういった点で比較すれば良いかを明確にするために、比較の目的を入力として与える。たとえば、「豚」と「牛」の「飼育のしやすさ」を比較する場合と、「食材としての調理法」を比較する場合では、比較の目的が異なる。

表 1: 比較の観点に応じた特徴の抽出の際に指定する条件 (A,B に比較対象の事柄を埋めて与える)

差分点	観点と A の特徴、観点と B の特徴が一致するように [MASK] に短文を当てはめて出力。観点は「比較の目的」に直接関係のある具体的な言葉で、最大数単語とする。特徴も目的を達成できる内容とし、可能であれば具体的な数値を用い、含意の広い回答は避ける。文はできるだけ短くし、接続詞を使って複数の項目を並べることは避ける。テンプレート以外の形式を用いることは禁止。
共通点	観点と A の特徴、観点と B の特徴が一致するように [MASK] に短文を当てはめて出力。観点は「目的」に沿って、最大で数単語とする。可能なら数値を用いるなど、具体的な内容を出し。複数の項目を並べることは避け、具体例を挙げることも避ける。テンプレート以外の形式を用いることは禁止。

表 2: 比較の観点に応じた特徴の抽出に用いるプロンプト (「**」に比較の目的, A,B に比較対象の事柄を埋めて与える)

差分点	「**」という比較の目的に直接関係のある、A と B の一般的で重要な違いを一つだけ挙げてください。観点は目的と異なる言葉を使ってください。
共通点	「**」という比較の目的に直接関係のある、A と B の一般的で重要な共通する特徴を一つだけ挙げてください。観点は目的と異なる言葉を使ってください。

3.3 比較の観点に応じた特徴の抽出

ユーザが入力した2つの事柄と比較の目的をもとに、比較の観点と、2つの事柄の差分点および共通点を抽出する。

具体的には、表1の条件と表2のプロンプトを用いて、ChatGPT に表3のフォーマットで一文で回答してもらうことで、比較の観点と2つの事柄の差分点および共通点を得る。

本節の処理は、ユーザが指定した数（デフォルト値は両者とも3）の差分点と共通点を得られるまで繰り返す。ただし繰り返しを行う際には、以前に出力された内容をプロンプトに含めることで、それらと異なる回答をしてもらう。また、3.4.1 項と 3.4.2 項の処理において、適切でないと判定された出力がある場合にも、再び本節の処理を実行して出力を追加する。

3.4 比較の観点に応じた特徴の正誤判定

3.4.1 特徴の正誤判定

前節で抽出された差分点と共通点の内容が正しいかどうかを確認する。ChatGPT が生成するテキストに

表 3: 差分点および共通点を抽出する ChatGPT の回答フォーマット (A,B に比較対象の事柄を埋めて与える。[MASK] 部分を埋めた一文を回答してもらう)

差分点	[MASK] の観点で、A は [MASK]、B は [MASK] という違いがあります。
共通点	[MASK] の観点で、A と B の共通点はどちらも [MASK] です。

表 4: 差分点および共通点の正誤判定を行うプロンプト (続くプロンプトで「観点」「A の特徴」「B の特徴」を与える)

差分点	数字のみ出力。与えられた内容が以下のどれにあたるか判定。0: 2 つの特徴が同じ / 0: 少なくとも一方の特徴が誤り・誤解を招く表現を含む / 1: 両方の特徴が一般的に正しい
共通点	数字のみ出力。与えられた内容が以下のどちらにあたるか判定。0: 特徴に誤り・誤解を招く表現を含む / 1: 一般的に正しい

は誤った内容が含まれる可能性があるため、それらを可能な限り排除することを目指す。

ChatGPT を用いて、差分点および共通点の正誤判定を行うプロンプトを表4に示す。「正しい」の判定基準においては、一部の例外を考慮しないようにするため、「一般的に正しい」という表現を用いている。例えば、「ボールペン」は書いたインクを消せないのが一般的だが、一部消すことができるインクのボールペンが存在する。そのため、多くの場合に正しいと考えられるものは正しいと判定させる。また、差分点において共通の特徴が出力されるケースがあり、その特徴が間違いとは言えない場合に、差分点としては不適切なため、「2 つの特徴が同じ」という判定結果をプロンプトに含めて除くようにしている。

この処理によって「一般的に正しい」と判定されなかった差分点と共通点は、対応する観点とともに出力から除くとともに、新たな差分点や共通点を追加するために、3.3 節の処理を再度実行する。

3.4.2 観点と特徴の対応判定

抽出された差分点と共通点が、ともに抽出された観点对応する関係にあるかを確認する。表5のプロンプトを ChatGPT に与えて、得られた差分点が観点对応しているかを確認する。共通点と観点の対応についても、同様のプロンプトを用いて確認する。3.3 節の回答フォーマットを用いて得た差分点と共通点は、基本的に観点に対応していることが期待されるが、誤りが生じるケースがあったため、本節の処理による確認を行う。

表 5: 観点と差分点の対応確認に用いるプロンプト（続くプロンプトで「観点」「A の特徴」「B の特徴」を与える）

数字のみ出力。ステップバイステップで、A の特徴が観点を説明するものか、B の特徴が観点を説明するものかを確認。両方が観点を説明するものである場合は 1、少なくとも一方が違う場合は 0 を出力。

表 6: 目的に対する重要度予測に用いるプロンプト（「**」に比較の目的を埋めて与える。続くプロンプトで差分点または共通点を与える）

数字のみ出力。与えられた内容が「**」という比較の目的に対してどれだけ重要か 5 段階で予想。A と B の違いをよく知る 100 人に尋ねたとき、最初にその違いを挙げる人数を考える。早く挙げるものほど一般により重要とする。

100 人中最初に挙げる人数の参考# 5:90 人以上/ 4:70-90 人/ 3:50-70 人/ 2:25-50 人/ 1:25 人以下

3.5 比較の観点に応じた特徴の重要度評価

抽出された差分点や共通点が、比較の目的に対してどれくらい重要と考えられるかを予測し、出力の際の順位づけに用いる。

具体的には、表 6 のプロンプトを ChatGPT に与えて、得られた観点の重要度を 5 段階評価で予測する。本研究においては、提示する差分点や共通点が、より多くの人に認知されているものほど重要と考えたため、100 人中の認知されている人数をもとに評価を行っている。

3.6 比較の観点に応じた特徴の分類

得られる差分点は、そのすべてがある観点（属性）の異なる属性値で表されるものばかりではなく、観点に対して当てはまる程度が数値的に表されるものが存在する。

数値的に大小関係を比較可能な差分点は、数直線上に表した方がよりユーザが理解しやすいと考えた。そこで、得られた差分点がある同一数直線上の値として表現が可能かどうかを、表 7 のプロンプトを用いて判定する。この判定の結果、数値的に表現が可能と判定された差分点について、具体的な数直線上の数値を、表 8 のプロンプトを用いて得る。与える数値は、差分点の中の 2 つの事柄の特徴のうち、比較の目的に対してより利点が大きいのと考えられる方を高く設定する。全ての差分点において、2 つの事柄の特徴の数値の合計を 1 に統一し、可視化を単純にする。

表 7: 差分点が数直線上の値として表現可能かを判定するプロンプト（続くプロンプトで「A の特徴」「B の特徴」を与える）

数字のみ出力。2 つの特徴が数字の大小に関する特徴である場合は 1、そうでない場合は 0 を出力。ただし、2 つの特徴の一方が「ある」、他方が「ない」を表す場合も 1 を出力。

表 8: 差分点の数直線上の値を求めるためのプロンプト（「**」に比較の目的を埋めて与える。続くプロンプトで「観点」「A の特徴」「B の特徴」を与える）

辞書形式で出力。観点と A の特徴の関連度、観点と B の特徴の関連度を、2 つの合計が 1 になるように設定。「**」という目的に対する利点の大きい方の関連度を高く設定し 0.1 から 0.9 の間の小数で表す。ただし、0.5 を使用することは禁止。なお、与えられた比較文が、一方が観点を持つ、他方が持たないという形式の場合のみ、持つ側を 1、持たない側を 0 とする。それ以外で 1 や 0 を使用することは禁止。

出力形式：“A”：関連度，“B”：関連度

3.7 2つの事柄の比較支援システムのインタフェース

前節までの処理によって得られた、比較の観点、差分点、共通点をインタフェース上に出力する。図 2 にインタフェースの表示例を示す。インタフェースは、3 つのウィンドウから構成されており、一番左のウィンドウは入力インタフェースを兼ねている。

左のインタフェースの上部に、比較したい 2 つの事柄と比較の目的、出力してほしい差分点と共通点の数を入力する。表示例においては、2 つの事柄として「自家用車」と「新幹線」、比較の目的を「家族旅行での移動手段を決める」として、差分点を 5 つ、共通点を 3 つ出力するように入力している。左のウィンドウには、3.3 節のテンプレートに沿って ChatGPT が回答した差分点と共通点を、中央のウィンドウには観点を、それぞれ重要度の順に出力する。

また右のウィンドウには、数値的に比較可能な観点による差分点と共通点を数直線形式で、それ以外の観点による差分点と共通点を表形式で出力する。

表形式の表示では、差分点の場合はそれぞれの特徴を赤、青の色付きで表示し、共通点の場合は観点を緑で表示する。

「自家用車」と「新幹線」の、「家族旅行での移動手段を決める」という目的に対する比較結果として、荷物の自由度や移動の快適さにおいて自家用車の方が多いという内容の出力が得られた。数直線による表示も踏まえて、自家用車の方が利便性がより高いと言える結果が得られた。



図 2: 2つの事柄の比較支援システムのインタフェース

4 2つの事柄の比較支援システムの評価実験

本章では、提案する2つの事柄の比較支援システムが、実際に2つの事柄の違いを比較するために役立てられるかを検証するためにに行った実験について述べる。

4.1 実験手順

表9に示す5つのトピックに関して、「比較の目的」についての相談を友人にされた状況を想定して、選択肢を比較した上で被験者が良いと思う選択肢を回答してもらう実験を行った。システムが出力する「違い」と「共通点」は、それぞれ3件とした。これは、主要な違いと共通点を出力することで、提案システムを用いて、要点を押さえた比較が行えることを確認するためである。実験は、16名の理系の大学生、大学院生に対して行い、提案システムを用いるグループとシステムを用いないグループとに8人ずつに分けて実施した。1つのトピックについて、提案システムを用いるグループの実験手順は以下の通りとした。

1. トピックについて「比較の目的」と「比較する選択肢」を提示する。
2. 「比較する選択肢」の中から、トピック「日本の都市」「外国語」では3つの選択肢、その他のトピックでは2つの選択肢を直感で選んでもらう。

3. 2つの選択肢を比較した提案システムの出力を提示して、2つの選択肢の「違い」と「共通点」をそれぞれ1個から3個の範囲で回答してもらう。
4. 3.の手順を、選んだ選択肢のすべての選択肢の組み合わせに対して行ってもらう。
5. 2.で選んだ選択肢に優先順位をつけ、自分が良いと思う選択肢1つを回答してもらう。

2.の手順により、被験者ごとに異なる選択肢の比較を促すことを意図した。これは、一般的にも多くの選択肢の中からある程度可能性がありそうな選択肢を絞ってから、その詳細を比較すると考えられたことによる。また、システムが提示する比較結果が固定されることで、恣意的な結果の提示にならないために行った。

4.の手順において、「違い」と「共通点」は、それぞれ「○○の点で、Aは△△、Bは××という違いがある。」「○○の点で、どちらも△△である。」という形式で回答してもらった。また、回答はシステムの出力を参考にしても良いし、被験者自身が持つ知識の中から回答しても良いこととした。これは、違いを考える主体はユーザにあり、システムはあくまでユーザの支援として用いられることを想定したことによる。

一方、提案システムを用いないグループの被験者には、上記の手順を提案システムを利用せずに行ってもらった。ただし提案システムの代わりとして、インターネットによるWebサイトの閲覧を可能とした。これは、一般的に比較を行う際には、インターネットを利用し

表 9: 実験に用いたトピックと比較する選択肢

トピック	比較の目的	比較する選択肢
日本の都市	生活スタイルに合う 場所に引っ越す	青森, 秋田, 盛岡, 福島, 山形, 仙台
スマホの OS	使用感が合う スマホを選択する	iOS, Android
携帯キャリア	生活に合う キャリアを選択する	softbank, docomo, au, 楽天モバイル
外国語	将来を見据えて 勉強する	フランス語, スペイン語, ドイツ語, イタリア語, 中国語, 韓国語
動画配信サービス	アニメを見るための サービスを選ぶ	Amazon Prime Video, Netflix, d アニメストア, U-NEXT

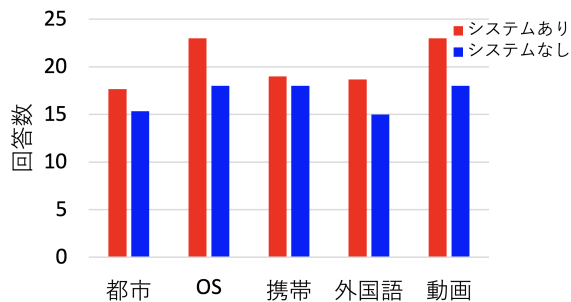


図 3: 比較 1 回あたりの「違い」の回答数 (被験者合計)

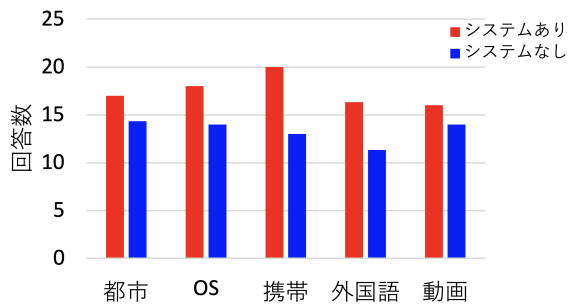


図 4: 比較 1 回あたりの「共通点」の回答数 (被験者合計)

て探索が行われることを想定したことによる。

また、提案システムを用いるグループの被験者には、回答時に提案システムの出力を参考にした回答には、チェックをしてもらうように指示を行った。

4.2 実験結果と考察

図 3 と図 4 に、比較 1 回あたりの「違い」と「共通点」の回答数の被験者合計を示す。結果として、すべてのトピックにおいて、提案システムを利用した比較の方が、「違い」と「共通点」の両方において、より多くの回答がなされていた。このことから、提案システムを利用することで、独力で比較することに比べて、より多くの違いと共通点の気づきにつながると考えられる。

表 10: システムを用いた被験者がシステムの出力を参考に回答した割合

トピック	違い	共通点
都市	94%	100%
OS	70%	94%
携帯	89%	95%
外国語	91%	96%
動画	83%	94%
平均	85%	96%

表 10 に、システムを用いた被験者がシステムの出力を参考に回答した割合を示す。全体的に高い割合で、システムの出力を参考に回答が行われていた。このことから、システムの出力が 2 つの事柄を比較するための違いの出力として、有効に用いられたことがわかる。

続いて ChatGPT によって、被験者の「違い」と「共通点」の回答が、比較の目的に対してどの程度役立てられるかを、表 11 の基準で 4 段階評価で点数をつけてもらった。図 5 と図 6 に、被験者の「違い」と「共通点」の ChatGPT による評価値を示す。

全体的に、提案システムを用いた回答の評価値の方が高い結果となった。提案システムの出力は、ChatGPT を用いて比較の目的に関係のある違いや共通点を挙げてもらっている。それを自己評価していることになるため、提案システムの回答を多く利用した被験者の回答の評価値が高くなることはある意味当然と考えられる。しかし、ChatGPT の知識の範囲は人間の知識の範囲よりも一般的に広く、ChatGPT を利用した回答が人間の独自の回答よりも評価値が高い結果となったことは、一般的な違いの比較においては、提案システムを用いたほうが、独力で比較することに比べてより効果的な比較が行える可能性を示唆しているとも言える。

図 5 において、トピック「携帯」と「動画」においては、評価値に大きな差は現れなかった。これは、「都市」「OS」「外国語」に比べると、普段から馴染みが深く、被験者の事前知識があったことが理由として考えられる。また、トピック「携帯」と「動画」の評価値は、他の「都市」「OS」「外国語」に比べると高い値となっており、評価値が低いトピックにおける比較ほど、シ

表 11: ChatGPT による回答の質的評価の基準

評価	基準
4	非常に役に立つ
3	ある程度役に立つ
2	あまり役に立たない
1	まったく役に立たない

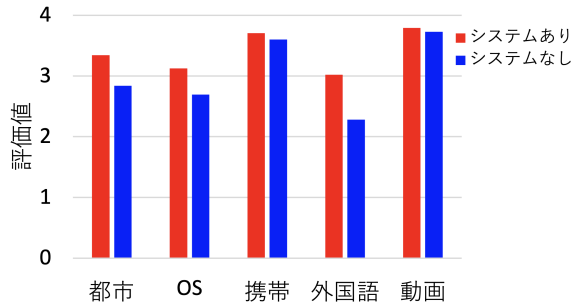


図 5: 「違い」の ChatGPT による評価値 (全回答平均)

システムの利用の有無での差が大きくなる傾向があった。

具体的には、評価値が高かったトピック「動画」においては、視聴可能なアニメの作品数やサービス料金などの、アニメ視聴に利用するサービスを決定する際に重要と考えられる内容の回答を、システムの有無に関わらず多くの被験者が行うことができていた。一方で、評価値が低かったトピック「外国語」においては、システム利用グループでは、ビジネスでの利用機会などの回答が見られたが、システム利用なしグループでは、文字の違いや、文法的な語順の違いなど、比較的狭い範囲での比較にとどまることが多かった。

これらのことから、普段からの馴染みがない事柄や、事前知識が多くない事柄を比較する際には、提案システムを利用した際の効果がより大きくなると考えられる。

表 12 に、システムを利用した被験者に、システムの出力の適切さを 5 段階 (5: 全体的に適切だった, 4: どちらかという適切だった, 3: どちらでもない, 2: どちらかという適切でなかった, 1: 全体的に適切でなかった) で回答してもらったアンケート結果の被験者平均値を示す。全体的に高い値が得られており、主観的にも違いの比較に役立てられる感覚を得ていたことがうかがえる。

違いの回答においては、トピック「携帯」の評価値が 3.9 で、最も低い値となっていた。表 13 に、トピック「携帯」において被験者が評価値 2 をつけた提案システムの差分点の出力を示す。違いの出力としては、1 つ目と 3 つ目が料金に関する話で類似しており、また 1 つ目と 2 つ目がプランという点での類似が見られ、違いの出力に大きな重複が見られたことが評価を下げた原因として考えられる。この例のように、観点の単語が異なっている、比較している内容が類似している例は他にも見られたため、観点だけでなく、内容も含

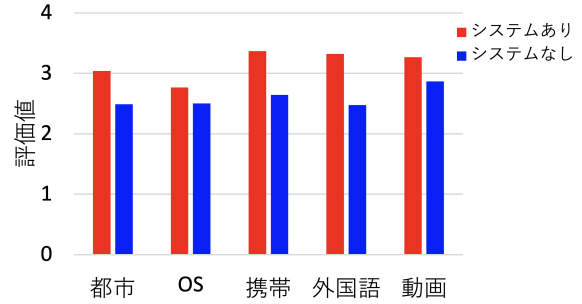


図 6: 「共通点」の ChatGPT による評価値 (全回答平均)

表 12: システムを利用した被験者に対するシステムの出力の適切さのアンケート結果 (5 段階評価の被験者平均)

トピック	違い	共通点
都市	4.1	4.0
OS	4.5	3.4
携帯	3.9	4.3
外国語	4.3	4.1
動画	4.5	3.9
平均	4.3	3.9

めて類似しない特徴が列挙できるように、システムを改善していきたい。

表 14 に、トピック「都市」における被験者の回答のカテゴリごとの回答数を示す。本トピックにおいては、各被験者は 3 回の比較を行い、各グループの被験者は 8 名のため、カテゴリごとの最大回答数は 24 件となる。なお、1 つのカテゴリに分類される複数の回答があった場合でも、1 件としてカウントしている。

トピック「都市」では、東北地方の都道府県の中から「生活スタイルに合う場所に引っ越す」という目的の比較をする必要があった。それに対し、冬の豪雪が想定されるため「気候」についての回答、ならびに生活における移動の足となる「交通」についての回答が、システムの利用の有無に関わらず多かった。しかし、システムを利用した被験者のほとんどが、システムの出力を参考に「気候」と「交通」に関する比較を回答したのに対して、システムを利用しなかった被験者においては、回答数がそこまで多くはならなかった。

また、「人口」や「政令指定都市」といったカテゴリは、システムを利用した被験者はほとんど回答せず、システムを利用しなかった被験者の回答には見られた。今回の実験においては、システムの「違い」の出力は 3 件までとしていたため、システムとして優先順位が低いカテゴリの回答は出力されず、システムを利用した被験者も回答しなかったことがうかがえる。

このような傾向は他のトピックにおいても同様に見られた。システムを利用した被験者の回答は、システ

表 13: トピック「携帯」において被験者が評価値 2 をつけた提案システムの差分点の出力（ハルシネーションの可能性を鑑みて，論文では会社名を伏せている）

料金プランの観点で、**は多様なプランがあり、**はシンプルなプランがあるという違いがあります。
データ使用量無制限の観点で、**は大容量プランがあり、**は無制限プランがあるという違いがあります。
料金体系の観点で、**は多様な家族割引があり、**は使った分だけ料金を支払うという違いがあります。

表 14: トピック「都市」における被験者の回答のカテゴリごとの回答数

カテゴリ	システムあり	システムなし
気候	20	12
交通	20	8
食文化	4	6
イベント	6	7
金銭	8	3
人口	1	8
政令指定都市	0	6
小計	59	50

ムの出力を参考に主要なカテゴリに関する比較を行い、システムを利用しなかった被験者においては、回答のカテゴリが多岐に渡る傾向が見られた。すなわち、主要な比較を忘れずに行っていく際には、システムの出力が被験者の助けとして用いられる可能性が高く、主要な比較を行った後に、提案システムの出力数を増やして比較していくことも考えられる。

5 おわりに

本論文では、入力された 2 つの事柄と比較の目的に対して、ChatGPT を用いて、比較の観点を用意した上で、2 つの事柄の差分点と共通点をユーザに提供するシステムを提案した。評価実験の結果、提案システムを用いることで、2 つの事柄の主要な違いと共通点をより多く挙げやすくなることを確認した。特にユーザが事前知識を持たない事柄の比較において、ユーザの知識を補う効果があることが示唆された。

今後の課題としては、より詳細な差分点と共通点を出力に含めて、より多岐に渡る内容の出力を行えるようにする点、3 つ以上の事柄の比較を一度に行えるようにする点、ChatGPT が十分に知識を持ち合わせていない事柄について、外部知識を与えた上で比較結果を出力できるようにする点などが挙げられる。

参考文献

[1] OpenAI: ChatGPT, (URL) <https://openai.com/chatgpt/> (2024 年 12 月 30 日確認)

- [2] Sisi Peng, Dan Qu, Hao Zhang, Wenlin Zhang, Shunhang Li : Using Knowledge Induction Strategies: Llms Can Do Better in Knowledge-Driven Dialogue Tasks, Available at SSRN 4670218, (2023)
- [3] 村上一彦, 瀬川修, 木村佳央: RAG を用いたヒューマンエラー事象の対策立案, 研究報告音声言語情報処理, 2024-SLP-154, No.18, pp.1 – 5, (2024)
- [4] Shicheng Xu, Liang Pang, Huawei Shen, Xueqi Cheng, Tat-Seng Chua : Search-in-the-Chain: Interactively Enhancing Large Language Models with Search for Knowledge-intensive Tasks, WWW '24: Proceedings of the ACM Web Conference 2024, pp.1362 – 1373, (2024)
- [5] 鈴木啓, 大内紀知: テキストマイニングによる学会の特徴分析, 経営情報学会 2016 年秋季全国研究発表大会要旨集, pp.79 – 82, (2016)
- [6] 吉岡真治, 神門典子, 関洋平: 複数国の新聞サイトを比較分析する NSContrast の実験的分析, 情報処理学会研究報告情報基礎とアクセス技術, Vol.2011-IFAT-103, No.2, pp.1 – 8, (2011)
- [7] 張建偉, 河合由起子, 熊本忠彦, 白石優旗, 田中克己: 多様な印象に基づくニュースサイト報道傾向分析システム, 知能と情報, Vol.25, No.1, pp.568 – 582, (2013)
- [8] 楠村幸貴, 土方嘉徳, 西田正吾: NTM-Agent: テキストマイニングによるネットオークションの商品比較支援, 第 15 回データ工学ワークショップ論文集 (DEWS2004), I-4-06, (2004)
- [9] P. Ziemba, J. Jankowski, J. Watrobski : Online Comparison System with Certain and Uncertain Criteria Based on Multi-criteria Decision Analysis Method, In proc. of the 9th International Conference on Computational Collective Intelligence, pp. 579 – 589, (2017)
- [10] 梅原武志, 武田英明: 大規模言語モデルを利用したパンデミック期の事業等のリスクの記述分析, 第 38 回人工知能学会全国大会論文集, 2I6-GS-10-04, (2024)
- [11] Z. Ren, Y. Su, X. Liu: ChatGPT-Powered Hierarchical Comparisons for Image Classification, In Proc. of the Advances in Neural Information Processing Systems 36, (2023)

特徴タグ分析を用いたアニメーションに関する メタデータ作成の提案

Metadata Building Method for Anime Clip Scenes using Feature Tag Analysis

Shan Junjie^{1*} 石井 智也¹ 安尾 萌² 西原 陽子¹
Junjie Shan¹ Tomoya Ishii¹ Megumi Yasuo² Yoko Nishihara¹

¹ 立命館大学情報理工学部

¹ College of Information Science and Engineering, Ritsumeikan University

² 立命館グローバル・イノベーション研究機構

² Ritsumeikan Global Innovation Research Organization, Ritsumeikan University

Abstract: 近年、大規模言語モデル (LLM) の発展に伴い、画像や動画の内容に対し自然言語により概要を生成し、それを用いた検索を行う手法が流行している。しかし、自然言語による記述は、内容と形式の一貫性を保つことが困難であり、LLM によって生成された概要記述で画像や動画の統一的な分析・検索を行うのは困難である。本研究では、アニメーションに対する特徴タグ分析を行い、メタデータを構築する手法を提案する。分割されたアニメのカットシーンに対し、フレーム単位ごとに特徴タグを抽出することにより、統一的なタグ記述のあるアニメーションのデータセットを構築した。本論文では、アニメ動画の入力からカットシーンの特徴タグを生成するまでの処理の流れを説明し、生成された特徴タグを用いて異なるジャンルのアニメーションを分析した。

1 はじめに

近年、大規模言語モデルの発展により、ビデオの検索に注目が寄せられている。LLM を使用してビデオコンテンツの説明文 (Caption) を生成し、ユーザの検索クエリとのテキスト類似度を計算するのは一般的なアプローチである [1-3]。また、マルチモーダル LLM の発展により、直接特徴画像の入力からビデオのコンテンツから該当するシーンの特定も可能になった [4]。

しかし、LLM により生成された自然言語の説明概要は、自然言語自身の柔軟性により、用語、文法、形式などで記述的な統一性を保つことは困難である。例えば同じ入力ビデオやプロンプトであっても、モデル推論上の影響により、出力される記述文が変わる可能性も残っている。そのため、LLM が生成した記述文を用いて、ビデオコンテンツの統一的な分析や柔軟なインデックスを作成するためのメタデータの構築が難しくなる。

この問題に対し、本研究では、入力画像の内容的な特徴を分析し、タグとして出力するモデルを利用することで、ビデオコンテンツのシーン画面に対するメタデータの構築手法を提案する。提案手法では、世界中

で人気がある日本のアニメ動画をビデオコンテンツとして、アニメーションのメタデータの構築を実装した。

本研究では、主に以下の三つの内容を達成した：

1. 入力されたアニメ動画から、自動的に各カットシーンの画面内容に対する特徴タグのメタデータを作成するシステムを実装した。
2. 異なるジャンルのアニメ動画を用意し、10 万件以上のカットシーンを含む特徴タグのデータセットを作成した。
3. 作成されたアニメーションのタグデータセットを用いて、ジャンル別に内容の特徴の分析を行った。

2 関連研究

ビデオや動画に特徴タグを付けて、内容分析や検索、推薦を行うシステムに関する課題は様々存在する。中村らの研究では、ニュースの動画シーンに事前に手動でタグ付けが行われたデータを分析し、シーンを選択する手法を提案した [5]。上畑らの研究では、ニコニコ動画に付けられた動画のジャンル等のタグを利用し、タグ

*連絡先：立命館大学情報理工学部
〒567-8570 大阪府茨木市岩倉町 2 - 150
E-mail: shan@fc.ritsumei.ac.jp

数やタグの種類を用いて、動画の多様性を分析した [6]. 村上らの研究でも、ニコニコ動画に関するタグの研究を行っており、視聴者が付与したタグを階層化し、タグ階層化データから関連動画を提示する応用の可能性を示した [7]. 関連動画を提示することで検索への応用も期待される.

これらの研究により、特徴タグがビデオの内容の一つの要約として、幅広い場面での応用可能性が示された. しかし、既存する特徴タグは主に沢山のユーザや専門家から手動で付けられ、ビデオ内容の広いジャンルや全体的な説明までしか反映できないものである. 本研究では、入力画像の内容を表す特徴タグを推定するモデルを利用し、自動的にアニメビデオの入力から各カットシーンの特徴タグを、より細部な画面内容を反映できるタグデータを作成するシステムを提案した.

一方、アニメ内容に関する特徴抽出やメタデータ作成の課題も多数存在する. 高山ら [8], 山田ら [9], 鈴木ら [10] はアニメのキャラクタを対象にして、色や位置などの視覚情報を利用する特徴抽出や類似検索の課題を行った. 高山らは肌の色やあごの形状などの特徴を用いて、アニメキャラクタの顔の特徴データセットを構築しており、山田らはグレースケール画像向きの顔の特徴抽出手法を提案した. 鈴木らはアニメキャラクタの顔パーツ検出手法に用いる学習データセットの作成案を提案した. アニメコンテンツに関するメタデータ付きのデータセットとしては、Manga109 [11] が挙げられる. 漫画作品をページ単位にして、「コマ」「キャラクタ」「テキスト」の三つの内容に対して人手で領域とラベルのアノテーションを行った. 品質の高いメタデータを構築した一方、大量なアノテーション作業やチェック管理にコストを要した. 本研究では、Web上に公開された画像をマルチラベルに分類するモデルを活用し、自動的にアニメ動画からカットシーン単位の特徴タグを付与する手法を提案する.

3 提案手法

3.1 システムの流れ

図1に提案したアニメ動画の入力からカットシーンの特徴タグを作成するまでの処理の流れを示す. 本研究では、アニメ動画のビデオデータからカットシーンを分割し、画像のマルチラベル分類モデルを用いてフレームごとに特徴タグを推定することにより、シーン画面の内容を表すタグデータを作成する.

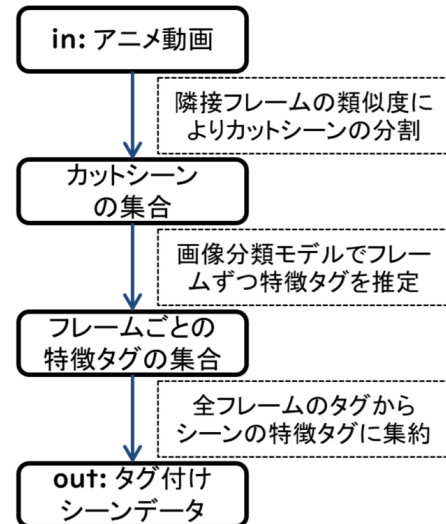


図1: アニメ動画からカットシーンの特徴タグを作成する流れ

3.2 カットシーンの分割

本研究では、動画構成の最小単位である「カット」を、特徴タグを付与するシーン対象として扱っている. 「カット」とは、一つの構図を表現し、同一オブジェクトが連続する画面であり、内容的には一貫性のある連続するフレームの集合である^{1 2}.

そのため、画面の内容的な連続性のあるカットシーンを分割する際に、隣接するフレームの類似度を測り、設定された閾値を超える（類似度が小さい）隣接フレームを前後に分けて、前のフレームまでの連続するフレーム集合を一つのカットシーンとして抽出する. 本研究では、アクション・スポーツ・日常系の三つのジャンルからそれぞれ10個のカットシーンに対する予備的な調査により、以下のシーン分割のルールを設定した.

1. 隣接フレームの類似度の測り方は、OpenCVのテンプレートマッチングにある「TM_CCOEFF_NORMED」方法³を利用する.
2. 類似度「TM_CCOEFF_NORMED」の閾値を0.6に設定し、類似度が0.6以下である隣接フレームを前後に分ける.
3. フレーム数が10枚以下のカットシーンが連続して出現した場合、再生時間の順番に合わせて、前からフレーム数10枚未満のカットシーンを30枚以上までに合併する.

以上のルールに従って分割されたカットシーンを、三つのジャンルからそれぞれランダムに1000件抽出し、

¹<https://kotatsu.info/2022-06-09-movie-terminology/>

²<https://orita-ani.net/scene-different-from-sequence/>

³https://docs.opencv.org/4.x/df/dfb/group_imgproc_object.html

表 1: 分析する特徴タグの情報とサンプル.

タグ種類	サンプルタグ	タグの説明
Character (キャラ関連のタグ, 合計 2748 個)	blue_eyes	画面上に青い目のキャラがいる
	blonde_hair	画面上に金髪のキャラがいる
	kimono	画面上に着物を着るキャラがいる
	hair_over_eyes	目が前髪に隠されるキャラがいる
	shouting	画面に大声を出そうとしているキャラがいる
Background (背景関連のタグ, 合計 394 個)	cloud	雲が背景にある
	cliff	背景に崖がある
	classroom	背景に教室の風景がある
	indoors	室内である背景
	in_water	キャラが水の中にいる
Composition (構図関連のタグ, 合計 122 個)	from_above	ハイアングルの構図、俯瞰
	from_outside	室外からの視点
	pov	一人称視点
	eye_focus	目のクローズアップ
	vanishing_point	消失点のある構図

合計 3000 件のカットシーンに対して、人手でカットシーンの分割精度を評価した結果、平均的な分割精度は 93.83% となった。

3.3 フレーム内容の特徴タグの推定

本研究では、Web 上に公開された訓練済みのアニメ画像向きの内容特徴を推定するモデル⁴を利用し、分割されたカットシーンのフレーム画面に対し一枚ずつ特徴タグの推定を行った。分析可能な特徴タグ数は合計で 6891 個であり、本研究では、その中に含まれる三種類の特徴タグを用いて、画面内容を表すメタデータを構築した。表 1 にフレーム画面の内容を表すタグとして抽出し、カットシーンに付与した三種類の特徴タグの情報とサンプルを示す。本研究では、キャラクタデザイン関連のタグ (Character)、画面背景関連のタグ (Background) と構図関連のタグ (Composition) の三種類から、アニメのカットシーンに対する内容のメタデータを作成している。キャラクタデザイン関連のタグは 2748 個、画面背景関連のタグは 394 個、構図関連のタグは 122 個を用意した⁵。

タグ推定モデルによりフレーム画面を一枚ずつ推定し、推定された内容タグの結果を合算し、一つのカット

⁴<https://github.com/KichangKim/DeepDanbooru>

⁵上記三種類のタグ以外に、オブジェクトを示すタグ (ペットボトル、ラーメン、ゲーム機など) や画像の表現を示すタグ (グレースケール画像、四コマ漫画など) も多数ある。今回はこれらのタグをカットシーンのメタデータ構築に使用しない。

```

CaptainTsubasa_005_cut0281.mp4:
[44, [0, 0, 44, 0, 0, ..., 0, 5, 0, 0, 0]]
カットシーンのID
Jujutsu_004_cut0049.mp4:
[55, [0, 3, 11, 0, 0, ..., 0, 0, 0, 0, 0]]

UmaMusume_S1_005_cut0065.mp4:
[226, [0, 0, 0, 0, 0, ..., 0, 70, 0, 0, 0]]
カットシーンの合計フレーム数 当該タグの出現したフレーム数
HeroAcademia_010_cut0329.mp4:
[36, [0, 22, 0, 0, 0, ..., 0, 0, 0, 0, 0]]
長さ6891の内容タグの位置に対応する格納ベクトル
HibikeEuphonium_S1_004_cut0334.mp4:
[56, [0, 56, 56, 0, 0, ..., 0, 0, 0, 0, 0]]

```

図 2: カットシーンの記録サンプル

シーンに対応するベクトルに格納する。図 2 にフレームずつ合算したカットシーンの記録サンプルを示す。一つのカットシーンに対して、シーンの ID、シーンのフレーム数と、フレームずつ合算した内容タグの格納ベクトルの三つの情報を記録する。

3.4 カットシーンの特徴タグ

同じカットシーンにしても、キャラクタやオブジェクトの移動、そしてカメラワークの動きにより、画面の内容も常に変わっている。一枚前のフレーム画面に存在する内容タグは、一枚後のフレーム画面から消える可能性もある。そのため本研究では、各カットシーンに対して、フレームごとに推定して合算された内容タグの格納ベクトルから、カットシーンの合計フレーム数の半分 (50%) 以上の出現頻度のあるタグを抽出し、当該カットシーンの特徴タグとしてシーンに付与する。

例えば、図 2 に示す一番下の記録のように、56 枚のフレームがあるカットシーンの場合、出現頻度は 28 回以上のタグ (=56*50%, 同じタグは 28 枚以上のフレーム画面から検出された) を、当該シーンの内容特徴を代表するタグと考え、カットシーンの特徴タグとして抽出する。出現頻度が 28 回未満の内容タグは、カットシーン全体の内容特徴を反映できないとみなし、特徴タグから削除する。

抽出されたカットシーンの特徴タグを、3.3 節に示す内容タグの格納ベクトルと同じ順に、長さ 6891 の One-Hot ベクトルの形で保存する (図 3)。

```

HeroAcademia_005_cut0073.mp4 [0, 0, 0, 0, 0, ..., 1, 0, 0, 0, 0]
MIX_S1_001_cut0093.mp4      [0, 1, 1, 0, 0, ..., 0, 1, 0, 0, 0]
DragonBall_107_cut0275.mp4  [0, 1, 1, 0, 0, ..., 1, 0, 0, 0, 0]

```

図 3: カットシーンの One-Hot 特徴タグベクトル

表 2: ジャンル別のカットシーンの統計.

ジャンル	カットシーン数	平均フレーム数
アクション	34673	127.97
スポーツ	31864	119.91
日常系	32601	120.54
合計	99138	122.94

表 3: 内容タグのカットシーンでの出現割合.

出現頻度	タグ数割合
10%未満	0.3934
10% ~ 20%未満	0.1116
20% ~ 30%未満	0.0723
30% ~ 40%未満	0.0552
40% ~ 50%未満	0.0446
50% ~ 60%未満	0.0401
60% ~ 70%未満	0.0352
70% ~ 80%未満	0.0320
80% ~ 90%未満	0.0337
90% ~ 100%未満	0.0556
=100%	0.1262

表 4: 「アクション」アニメでの種類別頻出タグ Top10.

タグ種類	タグ内容	出現シーン数 (割合)
Character	1girl	11228 (32.38%)
	1boy	8722 (25.15%)
	black_hair	8405 (24.24%)
	short_hair	5353 (15.44%)
	open_mouth	5344 (15.41%)
	long_hair	5191 (14.97%)
	smile	5134 (14.81%)
	looking_at_viewer	4759 (13.72%)
	brown_hair	4286 (12.36%)
	shirt	4119 (11.88%)
Background	sky	8205 (23.66%)
	outdoors	7813 (22.53%)
	cloud	6824 (19.68%)
	blue_sky	4264 (12.30%)
	cloudy_sky	4187 (12.07%)
	tree	3566 (10.28%)
	scenery	3360 (9.69%)
	building	2996 (8.64%)
	mountain	2066 (5.96%)
	city	1937 (5.59%)
Composition	male_focus	9034 (26.05%)
	close-up	3897 (11.24%)
	upper_body	3533 (10.19%)
	depth_of_field	2276 (6.56%)
	from_behind	567 (1.63%)
	lens_flare	567 (1.63%)
	from_side	567 (1.63%)
	eye_focus	479 (1.38%)
	gradient	469 (1.35%)
	monochrome	441 (1.27%)

4 タグ付けシーンデータの分析

本研究では、アクション・スポーツ・日常系の三つのアニメジャンルからそれぞれ 10 タイトル、各タイトルに対して 10 話の合計 300 話 (=3*10*10) のアニメ動画をデータセットとして、提案システムによりタグ付けされたアニメのシーンデータを作成した。表 2 に分割されたジャンル別のアニメシーンの統計情報を示す。提案手法で分割されたカットシーン数はアクション、スポーツ、日常系でそれぞれ 34673 本、31864 本と 32601 本となり、平均フレーム数も三ジャンルで同じ 120 枚程度であった。

4.1 内容タグの分布の統計

フレームごとに記録した各カットシーンの内容タグ (3.3 節) の結果を基にして、各内容タグのアニメのカットシーンでの出現分布を調査した。表 3 に全ての内容タグの、カットシーン内で出現したフレーム数の割合の統計結果を示す。

統計結果により、四割弱 (39.34%) の内容タグがカットシーン中で 10%未満のフレームにしか検出されなかった (フレーム数は 100 枚であるカットシーンの場合、当該タグは、10 枚未満のフレームにしか存在しなかった)。この結果から、カットシーンの画面に出た主な内容タグは、オブジェクトやカメラの動きなどにより、わずかなフレームからしか検出できなかったと考えられ、シー

ン全体を代表する特徴タグとして抽出する際に、タグの出現頻度によりフィルタリングする必要があると判断する。一方、12.62%の内容タグが、カットシーンの全フレーム画面に検出された (出現頻度=100%)。この結果から、分割されたカットシーンの画面に集中している内容もあり、タグ推定によりアニメシーンの特徴タグを作成するのは可能と考えられる。

4.2 ジャンル別の頻出特徴タグ

抽出されたカットシーンの特徴タグを 3.3 節に定義した三つのタグ種類に分けて、カットシーンでの頻出タグをアニメのジャンル別により調査した。表 4、表 5 と表 6 にそれぞれ「アクション」「スポーツ」「日常系」のアニメのカットシーンにおいて、「キャラクタ」「背景」「構図」の三種類から統計した頻出する特徴タグの上位 10 個の結果を示す。

「アクション」と「スポーツ」のアニメシーンにある「キャラクタ」に関する頻出特徴タグの上位 10 個の内容は全て一致し、「日常系」アニメの上位 10 個において唯一異なるのは「school_uniform」(学校制服)である (「1boy(一人の男がいる)」と入れ替える)。この結果から、アニメ動画はジャンルに関係なく、キャラクタのデザインに近い選択や構成があることが分かった。

表 5: 「スポーツ系」アニメでの種類別頻出タグ Top10.

タグ種類	タグ内容	出現シーン数 (割合)
Character	black_hair	12875 (40.41%)
	shirt	10695 (33.56%)
	lgirl	9281 (29.13%)
	lboy	9159 (28.74%)
	brown_hair	7626 (23.93%)
	open_mouth	7091 (22.25%)
	short_hair	6988 (21.93%)
	smile	5692 (17.86%)
	looking_at_viewer	5268 (16.53%)
Background	long_hair	4727 (14.83%)
	outdoors	8640 (27.12%)
	sky*	5917 (18.57%)
	blurry_background	4450 (13.97%)
	tree	4426 (13.89%)
	cloud	4074 (12.79%)
	indoors	3659 (11.48%)
	blue_sky*	3425 (10.75%)
	building	2997 (9.41%)
	scenery	2148 (6.74%)
Composition	cloudy_sky*	2037 (6.39%)
	male_focus	11258 (35.33%)
	upper_body	6032 (18.93%)
	depth_of_field	5205 (16.34%)
	close-up	2729 (8.56%)
	from_side	871 (2.73%)
	gradient	766 (2.40%)
	from_behind	747 (2.34%)
	lens_flare	496 (1.56%)
	halftone	420 (1.32%)
	letterboxed	416 (1.31%)

表 6: 「日常系」アニメでの種類別頻出タグ Top10.

タグ種類	タグ内容	出現シーン数 (割合)
Character	lgirl	13881 (42.58%)
	long_hair	13117 (40.23%)
	brown_hair	10490 (32.18%)
	short_hair	9327 (28.61%)
	open_mouth	7418 (22.75%)
	black_hair	7325 (22.47%)
	school_uniform	6800 (20.86%)
	smile	6071 (18.62%)
	shirt	5854 (17.96%)
Background	brown_eyes	4763 (14.61%)
	outdoors	6590 (20.21%)
	sky	5069 (15.55%)
	blurry_background	4308 (13.21%)
	indoors	4271 (13.10%)
	tree	4117 (12.63%)
	cloud	3584 (10.99%)
	blue_sky	3151 (9.67%)
	building	2799 (8.59%)
	scenery	2660 (8.16%)
Composition	window	2413 (7.40%)
	depth_of_field	5515 (16.92%)
	upper_body	3293 (10.10%)
	close-up	2440 (7.48%)
	male_focus	1681 (5.16%)
	gradient	674 (2.07%)
	from_side	661 (2.03%)
	from_behind	597 (1.83%)
	pov	513 (1.57%)
	film_grain	498 (1.53%)
	portrait	418 (1.28%)

「背景 (Background)」に関する特徴タグからも同じ傾向が見られる。収集されたアニメ動画はジャンルに関係なく、背景の部分に「outdoors」「scenery」「sky」「tree」など室外の景色に関する内容が多かった。この結果から、アニメシーンの背景や場所にも類似性があることが分かった。「アクション」アニメと違い、「スポーツ」と「日常系」のアニメシーンで頻出する背景に関する上位 10 個の特徴タグに「indoors (室内の風景)」の内容があり、「アクション」のジャンルにより、「スポーツ」と「日常系」のアニメには室内であるシーンが多いことが分かった。

「構図 (Composition)」では、「upper-body (上半身)」や「close-up」など、近距離から描写するシーンがジャンルに関係なく多数あり、「アクション」と「スポーツ」のアニメのほうが、男性のキャラクターを中心に描写する (「male.focus」) シーンは、「日常系」のアニメよりはるかに多くあった (26.05%, 35.33% VS 5.16%)。

5 おわりに

本研究では、画像内容の特徴タグを推定するモデルを利用し、アニメ動画のカットシーン向きのメタデータの構築手法を提案した。アクション・スポーツ・日常系の三つのアニメジャンルから 300 話のアニメ動画

を収集し、提案システムにより特徴タグが付与されたアニメのシーンデータの作成を試みた。作成したアニメシーンのタグデータを分析したことで、提案手法によりアニメシーンのメタデータの作成は可能であることが分かった。

最終的なメタデータの作成までに、まだ二つの課題が残っている。一つ目は、現在分類した三つのタグ種類の中に関連性のあるタグが存在する。例えば、表 5 の「Background」にある「sky」と「blue_sky」「cloudy_sky」のような上位・下位関係のあるタグが複数存在する。今後はタグの種類をより細かく分類し、タグ間の関係を付けた上で、内容タグのオントロジーにより、アニメシーンの特徴タグを付与する。二つ目は、各特徴タグと画面上の内容に対応する部分 (タグに関連する画面内容の位置情報) はまだ不明である。例えばシーンの背景に「sky」のタグがあるとしても、当該タグに関する画面内容の広さ、そしてタグ内容のシーンに対する重要度の判断はまだ達成されていない。今後は新たなタグ推定モデルを導入し、Attention の仕組みなどを用いることで、各タグの推定に関連するピクセルの抽出及び、関連する画面内容の位置情報もシーンデータに付与する。

参考文献

- [1] D. Han, E. Park, G. Lee, A. Lee, and N. Kwak: MERLIN: Multimodal Embedding Refinement via LLM-based Iterative Navigation for Text-Video Retrieval-Rerank Pipeline, *arXiv preprint*, arXiv:2407.12508 (2024).
- [2] Lei N, Cai J, Qian Y, Zheng Z, Han C, Liu Z, and Huang Q: A two-stage Chinese medical video retrieval framework with LLM, *CCF International Conference on Natural Language Processing and Chinese Computing*, Cham: Springer Nature Switzerland, (2023).
- [3] Xu J, Lan C, Xie W, Chen X, and Lu Y: Retrieval-based video language model for efficient long video question answering, *arXiv preprint*, arXiv:2312.04931 (2023).
- [4] Team Gemini: Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context, *arXiv preprint* arXiv:2403.05530 (2024).
- [5] 中村浩之, 小川祐樹, 諏訪博彦: 動画のタグ情報を利用したニュースシーンの選択手法の提案, 日本社会情報学会学会誌, Vol. 22, No. 2, pp. 27–38, (2011).
- [6] 上畑恭平, 伊東栄典: タグの類似度に着目した利用者投稿サイト動画の多様性分析, 電子情報通信学会技術研究報告: 信学技報, Vol. 115, No. 381, pp. 83–88, (2015).
- [7] 村上直至, 伊東栄典: 動画投稿サイトで付与された動画タグの階層化, 研究報告数理モデル化と問題解決 (MPS), Vol. 2010, No. 17, pp. 1–6, (2010).
- [8] 高山耕平, Henry Johan, 西田友是: 特徴抽出によるアニメキャラクターの顔認識, 情報処理学会第 74 回全国大会講演論文集, Vol. 2012, No. 1, pp. 673–674, (2012).
- [9] 山田太雅, 棟方渚, 小野哲雄: アニメキャラクターの顔における特徴量抽出手法の提案と実装, 情報処理学会研究報告エンタテインメントコンピューティング (EC), Vol. 2017, No. 6, pp. 1–6, (2014).
- [10] 鈴木絵理, 山内拓人, 鄭顕志: 色及び位置情報を用いたアニメキャラクターの目と口の抽出, 第 36 回人工知能学会全国大会論文集 (JSAI2022), Vol. 2022, p. 204-GS-7-03, 2022.
- [11] 藤本東, 小川徹, 山本和慶, 松井勇佑, 山崎俊彦, 相澤清晴: 学術用アノテーション付き漫画画像データセットの構築と解析, 映像情報メディア学会技術報告, Vol. 41, No. 5, pp. 35–40, (2017).

住宅内空気環境の改善を目的としたアドバイス提供手法に関する 基礎的検討

Preliminary Study on Information Delivery Method for Improving Indoor Air Quality in Residential Buildings

坂 裕奈¹ 服部 俊一^{2*} 三浦 輝久²
Yuna Saka¹, Shunichi Hattori², Teruhisa Miura²

¹ 青山学院大学大学院理工学研究科

¹ Graduate School of Science and Engineering, Aoyama Gakuin University

² 一般財団法人電力中央研究所

² Central Research Institute of Electric Power Industry

Abstract: Indoor Air Quality (IAQ) in residential buildings tends to decline due to factors such as insufficient ventilation and excessive use of combustion heating. However, it is often difficult for residents to perceive these changes. This paper proposes an information delivery method to encourage residents to improve IAQ of their houses. Clustering and time series data analysis were conducted using IAQ measurement data from 24 residences in Japan. The study also discusses more effective ways to extract the characteristics of each residence and determine the optimal timing for presenting advice to raise IAQ.

1 はじめに

住宅内の空気環境は、燃焼暖房や建築物などから排出される化学物質により汚染されることがある。特に燃焼暖房は、二酸化炭素 (CO₂)・窒素酸化物・揮発性有機化合物など数多くの汚染物質の発生源となっている [1]。このような空気環境の汚染を測る間接指標として、CO₂ 濃度が広く採用されている。世界各国で屋内・住宅内を対象とした CO₂ 濃度の基準値が定められており、日本では建築物衛生法により一定面積以上の建築物では 1,000ppm が上限と定められている [2]。CO₂ 濃度が 1,000ppm を超える環境は集中力の低下や喘息などをもたらす、室内で活動する人の快適性や健康を損ねることがわかっている [3, 4]。

空気環境を改善する方法として、換気のほか燃焼暖房の利用を控えてエアコンなど他の暖房手段に切り替えることが挙げられる。国内の住宅 24 軒に対し室内 CO₂ 濃度などを計測した調査 [5] では、多くの住宅において外気温低下による換気量減少や燃焼暖房の利用により CO₂ 濃度が上昇する傾向にあることが報告されている。ただし、必要以上の換気は多くの場合電力消費量の増加を伴い、冬季における燃焼暖房の利用控え

は室温低下による健康影響が懸念される。したがって、住宅内空気環境の改善には換気などの対策を適切なタイミングで、適度を実施することが重要と考える。しかし、一般的に空気環境汚染を人間が知覚することは困難であり、上述の調査 [5] では住宅内空気環境に関心がある居住者であっても適切な換気が行えていない場合もあることがわかっている。

そこで本稿では、住宅内で測定された CO₂ 濃度などのデータを活用し、空気環境を改善するための効果的なアドバイスを提供する手法を提案する。住宅内の省エネを目的としたアドバイス提供の事例は多く、家電単位などパーソナライズされた情報をリアルタイムに提供することで省エネ効果が向上することがわかっている [6]。住宅内空気環境においても、CO₂ 濃度上昇や暖房利用のタイミングなどを推定し、住宅ごとにパーソナライズされた情報をリアルタイムで提供できれば、空気環境改善による快適性向上や健康改善が期待できる。

以上に述べたようなアドバイス実現に必要な機能として、(i) 各住宅に対するプロファイリング、(ii) アドバイスを提示するタイミング推定、(iii) 居住者に提供するアドバイスの生成が挙げられる。このうち (i) については、住宅ごとの特性に応じたアドバイス提供のためのプロファイリングを実現するため、CO₂ 濃度の変動傾向に基づいた住宅のクラスタリングにより、各住

*連絡先：一般財団法人電力中央研究所
〒240-0196 神奈川県横浜須賀市長坂 2-6-1
E-mail: shattori@criepi.denken.or.jp

宅の空気環境の分類を試みる。(ii)については、CO₂濃度など住宅内で計測されたデータの変動傾向やその周期を分析することで、アドバイス提示の適切なタイミング推定を試みる。これらの検討結果に基づいて、提案手法の構築に必要な上記(i)から(iii)それぞれに求められる課題を考察する。

2 住宅内空気環境改善に向けたアドバイス提供手法に関する検討

1節で述べた通り、居住者が知覚できない住宅内空気環境に対する適切なアドバイスとして、パーソナライズされたリアルタイムな情報提供であることが必要と考える。これを踏まえて、本稿では効果的な空気環境改善アドバイスの実現には以下に示す3つの機能が必要と考える。

- (i) **各住宅のプロファイリング機能**：各住宅の属性・制約や居住者の生活スタイルなどの特性を推定することで、住宅およびその居住者固有の情報をアドバイスに反映する。
- (ii) **アドバイス提供タイミングの推定機能**：住宅内で計測されたデータから空気環境の悪化や居住者の行動を推定することで、アドバイスを適切なタイミングで提供する。
- (iii) **提示アドバイスの生成機能**：アドバイスの根拠となる各種情報を正確かつ簡潔にまとめることで、居住者の住宅内空気環境への理解や行動変容を促すアドバイスを作成する。

このうち、本節では(i)および(ii)を対象とした基礎的検討を実施する。検討に用いたデータセットの詳細については2.1節で、(i)、(ii)の検討についてはそれぞれ2.2、2.3節で述べる。

2.1 住宅内空気環境データセット

提案手法の検討にあたっては、先行研究[5]において収集された住宅内空気環境データセットを利用した。本データセットには関東地方を中心とした日本国内の住宅24軒を対象に計測されたデータが含まれており、本検証ではこのうち計測日時・CO₂濃度[ppm]・室温[°C]を利用する。なお、計測に用いたセンサの仕様により、CO₂濃度の計測範囲は360ppmから5,000ppmである。計測日時は各データが観測された日時を示している。本データセットの計測期間は2020年10月1日0時0分から2021年3月31日23時59分であり、計測粒度は1分である。

表 1: 3種類の期間でのCO₂濃度に基づく住宅のクラスタリングに対するクラスタ数別のSilhouette Score。

クラスタ数	期間		
	全期間	10月	12月
2	0.552	0.210	0.590
3	0.483	0.228	0.359
4	0.361	0.184	0.350
5	0.200	0.110	0.349
6	0.121	0.111	0.197
7	0.130	0.109	0.210
8	0.137	0.098	0.205
9	0.128	0.094	0.124
10	0.131	0.093	0.126

2.2 プロファイリング機能の検討

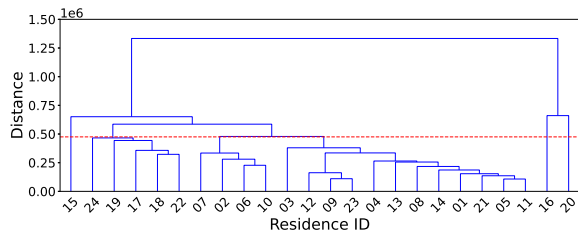
2節の冒頭(i)に示したプロファイリング機能の検討にあたり、CO₂濃度計測値に基づく住宅のクラスタリングを実施した。具体的には、特定の日時以前に各住宅で観測されたCO₂濃度について、この変動傾向が類似する住宅同士を複数のグループにまとめることで、各住宅の特徴を推定することを試みた。

事前分析として、住宅の分類に最適なクラスタ数を検討した。一定期間中に計測された各住宅のCO₂濃度に対して、Ward法に基づくクラスタ間距離測定方法を採用した階層的クラスタリングを実行した。クラスタリングに用いるCO₂濃度計測値の対象期間として、データセットに含まれる全期間、2020年10月、2020年12月の3通りを選択した。最適クラスタ数の決定には、クラスタ数を2から10に変動させた際のSilhouette Scoreおよびデンドログラムによる可視化結果を活用した。なお、Silhouette Score S は、住宅 $x \in X$ と同じクラスタに属する住宅との平均距離を a_x 、次に近いクラスタに属する住宅との平均距離を b_x とすると、以下の式(1)で表される。

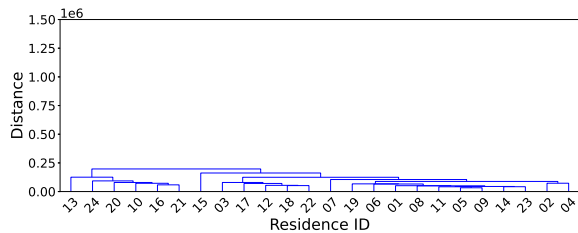
$$S = \frac{1}{|X|} \sum_{x \in X} \frac{b_x - a_x}{\max(a_x, b_x)} \quad (1)$$

S は-1から1の範囲を取り、1に近いほど同じクラスタ内のデータが互いに類似し、異なるクラスタとは明確に分離されていることを示す。

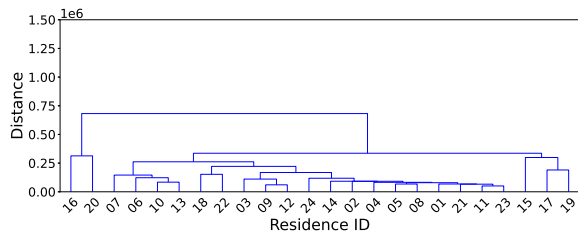
表1に、測定期間およびクラスタ数それぞれに対するSilhouette Scoreを示す。いずれの期間においてもクラスタ数が2または3の時にSilhouette Scoreが最大となり、クラスタ数が増えるにつれて下降する傾向が見られた。この結果から、数値上はクラスタ数が2または3が最善となるが、多様な住宅特性をプロファイリングするという目的に対してクラスタ数が過少であり、Silhouette Scoreも高いとはいえない。



(a) 全期間におけるデンドログラム。赤色破線は距離が 0.475×10^6 の位置に該当し、ここを基準にクラスタを 6 個に分割できる。



(b) 2020 年 10 月におけるデンドログラム。



(c) 2020 年 12 月におけるデンドログラム。

図 1: 3 種類の期間での CO₂ 濃度に基づく住宅のクラスタリング結果のデンドログラム。葉ノードは各住宅の ID、縦軸はクラスタ間距離を表す。

図 1 は、各期間におけるクラスタリング結果をデンドログラムで図示したものである。図 1(a) の全期間および図 1(c) の 2020 年 12 月では、一部住宅のみ他クラスタからの距離が遠く、他の住宅ではクラスタ間距離が短くなった。一方で、図 1(b) に示した 2020 年 10 月では全体的にクラスタ間距離が短くなった。また、図 1(a) 上の赤色破線で示した位置でクラスタを 6 個に分割すると、比較的距離の近い住宅同士でクラスタリングが行える。本稿ではクラスタ数を 6 に設定し、デンドログラムでの目視上距離が近い住宅を同じグループに割り当てることとする。

以上の結果を踏まえて、他クラスタに属する住宅と比較した CO₂ 濃度の変動傾向から居住者へのアドバイス提示を行う機能のプロトタイプを実装した。動作の流れとしては、任意の計測日時を指定し、これ以前の期間における各住宅の CO₂ 濃度計測値を用いて Ward 法に基づく階層的クラスタリングを行い、住宅を 6 クラスタに分類する。この結果に対して、指定したある住宅が属するクラスタに関する情報が出力される。

図 2 は、計測日時を 2020 年 12 月 9 日 12 時 0 分、住

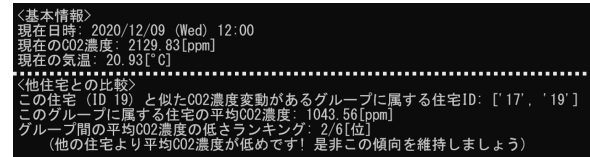


図 2: CO₂ 濃度の変動に基づく住宅クラスタリングを活用したアドバイス提示例。

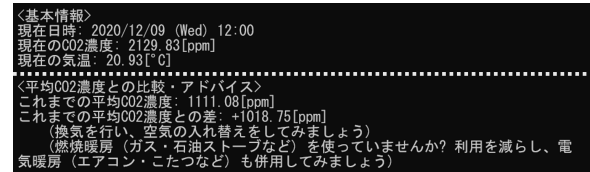


図 3: 計測日時・CO₂ 濃度・室温の計測データに基づくアドバイス提示例。

宅 ID を 19 と指定した際に出力された住宅内 CO₂ 濃度に関するクラスタ情報である。指定した住宅 ID と同じクラスタに属する住宅の ID、このクラスタに属する住宅のこれまでの平均 CO₂ 濃度などの情報が提示される。また、全クラスタにおけるクラスタ内の平均 CO₂ 濃度を順位として出力することで、他のクラスタに属する住宅と空気環境を比較できる。

2.3 アドバタイジング推定機能の検討

2 節冒頭 (ii) に示したアドバイス提供タイミングを推定する機能の実現に向けて、本節では計測値の活用および CO₂ 濃度変動周期の検出を試みた結果を述べる。

2.3.1 タイミング推定における計測値の活用

本データセットの計測日時・CO₂ 濃度・室温に基づいて、換気または暖房手段の変更をアドバイスすべきか判断する機能を実装した。具体的な挙動としては、指定した住宅および計測日時に対応する CO₂ 濃度を取得し、同時にこの住宅における指定日時（アドバイス提供タイミング）以前の CO₂ 濃度の平均値を算出する。これらの値を比較した際に指定日時の CO₂ 濃度が高い場合、相対的に空気環境が悪化しているとみなし、換気を促すアドバイスを出力する。この条件に加えて、指定日時が一般的な暖房の利用期間とされる 11 月から 3 月の間にあり、かつこの時の室温が 20°C を上回っていれば燃焼暖房は不要と判断し、暖房手段を切り替えるよう促すアドバイスを出力する。

日時を 2020 年 12 月 9 日 12 時 0 分、住宅 ID を 19 として指定した際に表示される空気環境情報およびアドバイスの例を図 3 に示す。図 4 に、この日時および直

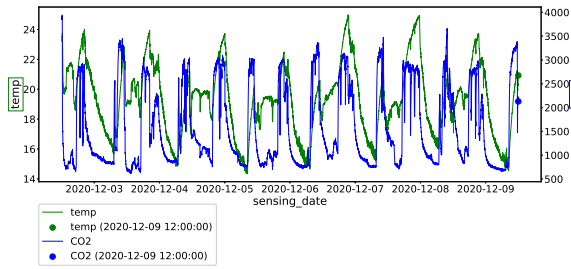


図 4: ID 19 の住宅での 2020 年 12 月 9 日 12 時 0 分とその直前 7 日間における CO₂ 濃度・室温. 青線は CO₂ 濃度, 緑線は室温を示し, 青点・緑点はそれぞれ 2020 年 12 月 9 日 12 時 0 分での CO₂ 濃度・室温を指す.

近 7 日間の CO₂ 濃度・室温を示す. 図 3 の出力例に示したように, 指定日時における CO₂ 濃度の計測値とこの日以前における CO₂ 濃度の平均値の差が算出され, 指定日時の CO₂ 濃度の方が高い場合は換気の実施を勧めるアドバイスが提示される. この例では, 指定日時が 11 月から 3 月の間かつ室温が 20°C より高かったため, 燃焼暖房の利用を控えエアコンなど燃焼を伴わない暖房手段への切り替えを勧めるアドバイスが出力されていることがわかる.

2.3.2 住宅内 CO₂ 濃度からの周期検出

空気環境の改善が必要なタイミングの推定にあたり, 本節では住宅内 CO₂ 濃度の増減から一定の周期性を検出することを試みる.

事前分析として, 各住宅における CO₂ 濃度の周期の有無および強弱の調査を実施した. 2.2 節で述べたクラスタリング結果のうち, 対象期間を全期間, クラスター数 6 とした場合の各クラスターから, それぞれ任意の住宅を 1 軒ずつ抽出した. 抽出された計 6 軒の住宅それぞれについて, 特定の 7 日間に計測された CO₂ 濃度に対して実数入力向けの 1 次元離散 Fourier 変換を適用した. 適用する 7 日間としては, 2020 年 10 月 29 日から 2020 年 11 月 4 日, および 2020 年 12 月 24 日から 2020 年 12 月 30 日を指定した. Fourier 変換のサンプリングレートは 1 日 (= 1,440 分) とした.

表 2 に, 各期間における各住宅の CO₂ 濃度に対して Fourier 変換を行った結果から, 上位 5 個のサンプル周期を示す. なお, サンプル周期が 1.000 未満のものは順位から除外した. 表 2(a) および (b) に示したいずれの期間においても, 下線で示した通りほぼすべての住宅にてサンプル周期 1.000 および 2.000 に対し Fourier 変換結果が大きくなる傾向が見られた. Fourier 変換のサンプリングレートが 1 日であることから, 住宅内 CO₂ 濃度に 1 日および半日間の周期が強く表れる傾向にあることを示している.

表 2: 各時期における各住宅の CO₂ 濃度に対する Fourier 変換結果上位 5 個のサンプル周期. サンプル周期 1.000 未満のものは順位に含まれていない.

(a) 2020 年 10 月 29 日 ~ 2020 年 11 月 4 日

順位	住宅 ID					
	01	02	15	16	19	20
1	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	<u>1.000</u>	<u>2.000</u>	<u>1.000</u>
2	2.571	<u>2.000</u>	2.286	<u>2.000</u>	2.143	<u>2.000</u>
3	<u>2.000</u>	4.000	2.571	1.286	1.714	4.000
4	3.143	1.857	3.714	3.000	4.000	3.429
5	2.143	2.143	5.571	3.714	1.286	6.000

(b) 2020 年 12 月 24 日 ~ 2020 年 12 月 30 日

順位	住宅 ID					
	01	02	15	16	19	20
1	<u>1.000</u>	<u>2.000</u>	<u>1.000</u>	<u>2.000</u>	<u>1.000</u>	<u>1.000</u>
2	1.143	1.286	2.857	<u>1.000</u>	<u>2.000</u>	<u>2.000</u>
3	<u>2.000</u>	1.714	2.429	3.000	1.143	4.000
4	2.571	<u>1.000</u>	2.286	1.571	3.000	1.857
5	3.000	3.286	5.571	1.429	2.714	6.000

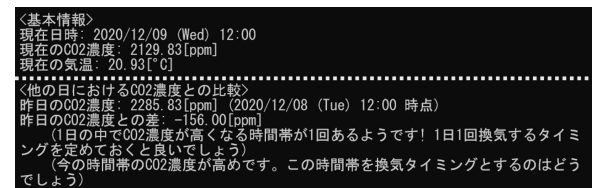


図 5: 直近 7 日間における住宅内 CO₂ 濃度から検出した周期の情報を活用したアドバイス提示例.

この結果を踏まえ, 直近の住宅内 CO₂ 濃度から特定の周期を検出し, これに基づきアドバイスを提示するタイミングを推定する機能を実装した. まず指定した住宅について, 指定計測日時から 7 日前までの CO₂ 濃度に対して上述の Fourier 変換を実行する. この変換結果の上位 10 個のサンプル周期に 1.000 または 2.000 が含まれる場合, この住宅の直近の CO₂ 濃度変動に 1 日または半日間の周期があるとみなし, アドバイス提示のタイミング推定に活用する. これに加えて, 2.3.1 節で述べた条件により CO₂ 濃度が相対的に高いと判断できる場合, この時間帯に空気環境が汚染される傾向にあると判断し, 併せてアドバイス内容に反映し空気環境の改善を促すことが可能になる.

図 5 に, この機能による出力例として日時を 2020 年 12 月 9 日 12 時 0 分, 住宅 ID を 19 と設定した際に提示されたアドバイスの例を示す. 図 4 に示した直近 7 日間の CO₂ 濃度分布からもわかるように, この例では 1 日間の周期が強いと推定されており, 指定日時から 1 日前の CO₂ 濃度およびこの値の指定日時の CO₂ 濃度

との差が提示された。加えて、図3でも示した例と同様に、この日時のCO₂濃度は平均CO₂濃度よりも高いため、指定日時の時間帯を換気タイミングとして設定することを勧めるアドバイスが出力されている。なお、半日間の周期が強いと判断された場合は12時間前のCO₂濃度に基づく情報が表示され、1日・半日間の周期が共に強いと推定された場合は1日前および12時間前のCO₂濃度に基づくアドバイスが提示される。

3 考察

前節では提案手法に関する基礎的検討として、プロファイリング機能およびアドバイス提供タイミングの推定機能について分析およびプロトタイプ構築を行った結果について述べた。この結果を踏まえて、提案手法構築に求められる要素について本節で考察する。

3.1 プロファイリング機能

2.2節で述べた検討により、室内空気環境が自宅と類似する住宅の有無やその数をクラスタリングにより分類することで、空気環境改善のアドバイスに活用できる可能性が示された。自宅が属するクラスターのCO₂濃度平均値を他クラスター値と比較し、その順位を把握することで、他の住宅と比較した際の空気環境の良し悪しを定量的に評価できると考える。この情報に基づくアドバイス提供は居住者の行動変容を促し、空気環境の自発的な改善が期待できる。さらに、居住者が自宅内の空気環境への理解を深めるだけでなく、他住宅と比較したCO₂濃度傾向を本手法により把握できれば、各住宅の特性や制約に応じた空気環境改善アドバイスが実現できると考える。

一方、本稿で行ったクラスタリングは、各住宅のCO₂濃度などの特徴を十分に反映できているとはいえず、空気環境改善に繋がるアドバイスに必要な情報をプロファイリングするという目的に対して課題が残る結果といえる。また、2.2節で行った階層的クラスタリングでは全住宅におけるCO₂濃度の計測データを必要とするため、新たに住宅を追加した場合には再度クラスタリングが必要になるといった課題も挙げられる。

これらの課題に対して、CO₂濃度以外の特徴量も用いたクラスタリングが必要と考える。本稿では各住宅のCO₂濃度のみを用いたクラスタリングを実施したが、たとえば利用する暖房手段や住居形態、世帯構成などの属性情報をアンケートにより収集してクラスタリングに反映できれば、各住宅の特性や制約を考慮した分類が実現できる。その他、深層学習系などの計算時間・計算量を低減できるようなクラスタリング手法の選定も改善にあたって効果的と考える。

3.2 アドバイスタイミング推定機能

2.3.1節で述べた計測値の活用については、各住宅で計測されたCO₂濃度分布を踏まえてアドバイスを提供することで、住宅それぞれで異なる空気環境の変化を考慮したタイミングでの情報提供が行えることを示した。CO₂濃度に加えて室温も活用することで、燃烧暖房の利用に関するアドバイスを適切なタイミングで提供できることがわかった。また、2.3.2節で述べた住宅内CO₂濃度の周期検出においては、CO₂濃度が1日または半日間の変動周期を持つ住宅を特定することで、換気などの空気環境改善行動を実施すべきタイミングを居住者が理解しやすい内容で提示できることを示した。

一方、本機能の検討で明らかになった課題として、アドバイス提供タイミングの判断基準となるしきい値の設定方法が挙げられる。本稿では直近一定期間内におけるCO₂濃度の平均値を基準とし、これを超えた場合に換気を促すアドバイスを提示している。この平均値がその住宅における標準的な空気環境を反映しているとは限らず、たとえば居住者の在宅割合が高い住宅と低い住宅とでは、同じ平均値であってもそれが意味するものは異なる。したがって、居住者の生活パターンやCO₂濃度分布を踏まえた基準値の算出が必要と考える。

加えて、暖房手段などの属性情報を考慮せずタイミング推定やアドバイス提供を行っていることも課題である。本稿では住宅内のCO₂濃度・温度のみを用いて推定していることから、燃烧暖房を所持していない住宅へ燃烧暖房に関するアドバイスが提供される可能性がある。この課題に対して、前節で述べたように属性情報をアンケートで収集するだけでなく、計測データから属性を推定するアプローチもユーザへの負担低減という観点で効果的と考える。先行研究[5, 7]では、CO₂濃度の急激な上昇や外気温との関連から暖房手段やその利用開始時期を推定できる可能性が示唆されている。一般的に燃烧暖房の利用は空気環境に与える影響が最も大きい行動であり、それに関するアドバイスは空気環境の改善余地も同様に大きい。このような属性情報を計測データから推定できれば、より効果的なアドバイスを適切なタイミングで、居住者への負担なく提供可能と考える。

3.3 アドバイス生成機能

本稿では、2節で述べたようにプロファイリング機能およびアドバイス提供タイミングの推定機能について検討を行った。しかし、居住者へアドバイスを提供するという提案手法の目的にはアドバイス生成機能も不可欠であることから、本機能の開発のために求められる課題について本節で考察する。

本稿では計測データのみを用いたアドバイスを提供したが、3.1 および 3.2 節で述べたように、居住者の実態に即したアドバイス実現には、住宅構造や生活パターンなどの属性情報の把握も不可欠と考える。そのためには、前述のように居住者へのアンケートや計測データからの推定による属性情報の収集が必要であり、特に後者においては電力消費量や照度・騒音など、様々な計測データの収集・活用が効果的と思われる。

また、2 節で構築したプロトタイプでは定型的なアドバイスが固定の条件に基づいて提供されており、温熱快適性や省エネに関する居住者の関心や選好を考慮していないことも課題といえる。この課題に対しては、生成 AI を活用した情報収集およびアドバイス生成が効果的と考える。生成 AI との対話を通じて、前述の関心や選好に加えて特定の時間帯やコンテキストに対する居住者からのフィードバックを獲得できれば、居住者の特性を踏まえたより効果的なアドバイス生成が期待できる。

生成したアドバイスを提供するためのインタフェースについても、今後の検討が望まれる。市販製品ではスマートフォンによる通知が一般的だが、ユーザは多くの通知に曝されているためアドバイスに気付かないなどの問題も懸念される。住宅内の省エネを対象とした調査 [8] において、リビングに設置した専用ディスプレイからの情報提供が日常的に閲覧されており、省エネ意識の醸成や教育に好影響を与える可能性が示唆されている。このようなインタフェースを用いることで、居住者が見落とすことなく、また負担を増大させないようなアドバイス提供・行動変容が実現可能と考える。

4 おわりに

本稿では、住宅内空気環境の改善を目的として、CO₂ 濃度などの計測データに基づくアドバイスを提供する手法を提案し、基礎的検討を行った結果について述べた。提案手法に求められる要素として、(i) 各住宅に対するプロファイリング、(ii) アドバイスを提示するタイミング推定、(iii) 居住者に提供するアドバイスの生成という 3 つの機能を挙げた。このうち (i) および (ii) については、CO₂ 濃度に基づく住宅のクラスタリングや CO₂ 濃度・室温を組み合わせたアドバイス提示を行うプロトタイプを構築した。その上で、本手法の実現に向けて (i) から (iii) それぞれについて求められる課題について考察・整理した。

今後は、本稿で示した課題に対しての更なる検討が望まれる。住宅内空気環境に関する調査は数多く実施されてきた一方で、その改善を目指した事例は少ない。省エネ分野などの事例を参考としつつ、居住者の意識・関心の向上やそれに伴う生活の質の改善に繋がる手法

構築が必要である。

参考文献

- [1] 野崎淳夫, 成田泰章, 二科妃里, 一條佑介, 山下祐希: 開放型石油暖房器具使用時の室内空気汚染に関する研究, 室内環境, Vol. 18, No. 1, pp. 33–44 (2015).
- [2] 市川勇, 松村年郎: 室内空気環境基準の現状, 衛生化学, Vol. 43, No. 3, pp. 162–173 (1997).
- [3] Allen, J. G., MacNaughton, P., Satish, U., Santanam, S., Vallarino, J., and Spengler, J. D.: Associations of Cognitive Function Scores with Carbon Dioxide, Ventilation, and Volatile Organic Compound Exposures in Office Workers: A Controlled Exposure Study of Green and Conventional Office Environments, Environmental Health Perspectives, Vol. 124, No. 6, pp. 805–812 (2016).
- [4] Simoni, M., Annesi-Maesano, I., Sigsgaard, T., Norback, D., Wieslander, G., Nystad, W., Canciani, M., Sestini, P., and Viegi, G.: School air quality related to dry cough, rhinitis and nasal patency in children, European Respiratory Journal, Vol. 35, No. 4, pp. 742–749 (2010).
- [5] Hattori, S., Iwamatsu, T., Miura, T., Tsutsumi, F., and Tanaka, N.: Investigation of Indoor Air Quality in Residential Buildings by Measuring CO₂ Concentration and a Questionnaire Survey, Sensors, Vol. 22, No. 19, Article 7331 (2022).
- [6] Ehrhardt-Martinez, K., Donnelly, K. A., and Laitner, J. A.: Advanced Metering Initiatives and Residential Feedback Programs: A Meta-Review for Household Electricity-Saving Opportunities, American Council for an Energy-Efficient Economy, (2010).
- [7] 服部俊一, 岩松俊哉, 三浦輝久, 堤富士雄, 田中伸幸: 住宅内空気環境の改善に向けた可視化・インタラクション手法の活用に関する検討, 第 26 回インタラクティブ情報アクセスと可視化マイニング研究会, pp. 29–34 (2021).
- [8] 服部俊一, 三浦輝久, 市川玲子, 澤井大樹: IoT センサを用いた家庭内の行動観察, 電気学会論文誌 C, Vol. 142, No. 2, pp. 206–215 (2022).

手技療法指導における動作指示の違いが指圧動作に及ぼす影響

Effects of Different Motion Instructions on Finger Pressure in Manual Therapy Training

田中 瑠慧 ^{✉,*}
Ryusei Tanaka

東岡 秀樹 [‡]
Hideki Higashioka

松下 光範 [‡]
Mitsunori Matsushita

[†] 関西大学大学院総合情報学研究科
Graduate School of Informatics, Kansai University

[‡] 関西大学総合情報学部
Faculty of Informatics, Kansai University

Abstract: 手技療法の指導では動作指示語が用いられているが、指導者が学習者に対して意図した動作を適切に伝えることは容易ではない。そこで本研究では、任意の動作に対して適切な動作指示語を明らかにし、手技療法指導の質向上に寄与することを目指す。本稿ではその端緒として、動作指示語が指圧動作に及ぼす影響を明らかにするため、15種類の動作指示語に基づく専門家3名の指圧動作からLSTMを用いて圧力・角度を推定し、分析を行った。

1 はじめに

急速な高齢化の進行の影響により、怪我や疲労に伴う身体的不調の回復・緩和を行う専門家（e.g., あん摩マッサージ指圧師、柔道整復師、理学療法士）の需要は高まっており、従事者の数も増加傾向にある。一方で、効率的な指導が十分に行われていないため、養成の質の低下^{*1}や施術の質の低下 [1] が懸念されており、指導方法の改善が必要である [2]。専門家の治療方法の一つとして、あん摩、マッサージ、指圧といった手技を用いた治療（以下、手技療法と記す）が挙げられる。手技療法の指導において、手技療法動作（e.g., 指圧、なでる、もむ、叩く）は、動作指示語（e.g., 強く押す、体の中心方向に向かって押す）を用いて伝達されていることが多い。この際、手技療法動作に対する力加減や角度は、指導者の身体感覚に依存するため、意図した手技療法動作を学習者に対して適切に伝え難いという懸念がある。そこで、本研究では手技療法における手技療法動作とその動作指示語の対応関係を明らかにすることで、指導者が学習者に対して任意の動作を伝達する際に、それに対応する適切な動

作指示語を選択し指導できるようにすることを目指す。

動作と動作指示語の関係を分析した試みとして、海老名ら [3] は動作指示語が触診における指圧動作に与える影響を明らかにした。また、動作指示語が「弱く」「普通の力で」「強く」と異なる強さの指圧を示す表現であるため、類似した強さの指圧の表現（e.g., 弱めに押す、少ない刺激で押す）の比較が行われていない。さらに、実験で活用した表現と手技療法で活用されている表現が異なる可能性がある。本稿では、田中らが実装した手技療法動作定量化システム [4] を活用し、定量化された手技療法動作と動作指示語の対応関係を明らかにする。システムは接触抵抗式感圧センサアレイで取得したデータから、Long Short Term Memory (LSTM) により手技療法動作時の圧力と角度を推定可能である。これを活用することで、手技療法動作に対する圧力や角度と動作指示語の対応関係の分析を試みる。これにより、これまで言語表現では正確に伝達できなかった情報を、手技療法動作定量化システムで得られる数値情報で補完することで、より正確な伝達が可能になることが期待される。

2 関連研究

本章では言語表現の特徴の定量的分析に関する研究として、動作指示語と定量化された動作の対応関係を分析した研究を概説する。

海老名らは協力者が触診を代行し計測値を遠隔地の医

* 連絡先：関西大学大学院総合情報学研究科
〒569-1095 大阪府高槻市霊仙寺町 2-1-1
E-mail: k290993@kansai-u.ac.jp

^{*1} <https://www.mhlw.go.jp/content/10801000/000499148.pdf>
(2024/2/27 確認)

表 1: 選定した手技療法動作指示語

分類	手技療法動作指示語
力の強さに関する語	強めに押す 弱めに押す 軽い刺激で押す 少ない刺激で押す 体重 3kg くらいかけて押す 柔らかく押す
力の角度に関する語	面に向かって 15 度くらいの角度で押す 真っ直ぐ上から押す 体の中心方向に向かって押す 頭方向に向かって押す 外側方向に向かって押す
圧力変化に関する語	柔らかく押す グーッと押す じんわり圧をかけて押す 3～5 秒かけてじっくり押す 3 秒かけてゆっくり押し、3 秒かけて指の力を抜きながら戻す
時間に関する語	3～5 秒かけてじっくり押す 3 秒かけてゆっくり押し、3 秒かけて指の力を抜きながら戻す

師に伝送する形態の遠隔触診の実現を目指し、その端緒として協力者が遠隔の医師からの指示通りの強さで対象物を指圧できるかを参加者 8 名による実験を行った。3 段階の程度副詞指示下（「弱く」/「普通の力で」/「強く」）での指圧力の個人差や指示間の力の比や、目標指圧力と実際の指圧力を定量的に提示した際の誤差について検証を行った [3]。その結果、指圧には標準偏差が 0.43kgf と一定の個人差がある一方、指示間の力の比は一定であることが明らかになった。また、定量的指示下ではすべての区分で平均二乗パーセント誤差の平方根が 10% 以下の高い精度で目標値を再現できることが明らかになった。

北尾らは、言語的な動作指示の違いが跳躍パフォーマンスに与える影響を、定量的に評価するための実験を行った [5]。動作指示語（1. 大きくジャンプ、2. 膝を深く曲げてジャンプ、3. 高くジャンプ、4. 最大限ジャンプ、5. 天井に触れるようにジャンプ）が跳躍に与える影響を Kinect を用いて分析した。その結果、ジャンプの高さは $5 > 4 > 3 > 2 > 1$ の順に高いことが明らかになった。また、5 のような天井などの身体の外環境を目標とする指示が高いジャンプを促すことに有効であることが示された。一方で、2 のように特定の身体部位に焦点を当てた指示では、指示された部位への注意が集中することで他の部位への注意が疎かになり、動作開始のタイミングが遅れる可能性が示唆された。また、4 のような動作イメージが具体的ではない指示は動作の再現性が低いことが明らかになった。

金田らは、健常男性 8 名を対象に、平行棒を使用して立つことを伝える理学療法士の 3 つの指示（free: 対象者の任意で平行棒を使用して立つ、push: 平行棒を押して立つ、pull: 平行棒を引いて立つ）が下肢関節モーメントに与える影響を明らかにした [6]。下肢関節モーメントは 3 次元動作解析装置を用いた動作分析手法によって、膝部離床時の股関節、膝関節、足関節の各モーメントを測定した。Push の指示では、膝関節伸展モーメントが小さく、Pull の指示では大きくなる傾向が明らかになった。また、平行棒を押した場合、平行棒には上向きの反力が発生し、床反力ベクトルの鉛直成分が小さくなるため、膝関節伸展モーメントが小さくなることが示された。これらの結果から、患者が立ち上がる際に理学療法士が「平行棒を押す」ように指示を出すことは下肢関節への負担を軽減し、安定した立ち上がりを誘発できる可能性が示唆された。

このように、動作指示語と定量化された動作の対応関係を分析した研究として、触診における指圧動作やジャンプ動作、立ち上がり動作を対象としたものが挙げられる。一方で、これまでの研究で扱われてきた動作指示とは異なり、手技療法動作指示は力の強弱、角度、圧力変化、時間といった複数の要素に影響を及ぼすため、従来の分析手法では不十分である。そこで、本研究では、手技療法動作定量化システムによって定量化された手技療法動作に対する 4 つの特徴（力の強弱、角度、圧力変化、時間）と手技療法における動作指示語との関係を明らかにすることを試みる。

表 2: 実験参加者の特性

	職業	経験年数	性別
1	あん摩マッサージ指圧師	2 年	男性
2	柔道整復師	18 年	男性
3	理学療法士	3 年	女性

3 実験

手技療法動作は指圧、なでる、もむ、叩くなどに大別されるが、本研究では手技療法の基本動作である指圧動作（押し込み）に着目し、分析を行った。本章では、手技療法に関する動作指示語（以下、手技療法動作指示語と記す）から、対象者が想起した指圧動作に対する特徴（力の強さ、力の角度、圧力変化、時間）の定量的分析について説明する。

3.1 手技療法動作指示語の選定

分析にあたり、力の強さ、力の角度、圧力変化、指圧時間に関連する指示を表す 15 個の手技療法動作指示語を選定した。YouTube などの Web サイトで用いられている手技療法動作指示語を参考に 28 種類の言語表現を抽出した。柔道整復師 1 名の協力を得て 15 種類に絞り込んだ（表 1 参照）。この際、28 種類の指示語のうち、「グッと押す」「ぎゅっと押す」など、被施術者の筋性防御を誘発する、手技療法動作として適切でないと考えられる 4 種類の表現を除外した。また、3.2 節で後述するように、実験では指圧動作が被施術者の特性（e.g., 筋力量、筋肉の疲労度、筋肉の硬結）による影響を受けないよう、マッサージ練習用のソフトビニール製の上半身モデルに対して指圧動作を行う。このモデルでは筋肉や骨の位置を正確に把握することが困難であるため、「横から背骨と筋肉をサンドイッチする感じに押す」や「脊柱起立筋に向かって真っ直ぐ押す」など、特定の筋肉や骨の部位を示す 9 種類の表現を除外した。

3.2 実験手順

本実験では、田中らの実装した手技療法定量化システム [4] を用いて、手技療法動作を定量化する。このシステムは、256 点の圧力を感知できる接触抵抗式感圧センサアレイで取得したデータに LSTM を適用することで、力の強さと角度（x 軸、y 軸）を推定可能である。センサアレイの中心部を加圧した際の推定精度は、平均二乗誤差（MSE）において、圧力値 5.19、x 軸角度 8.30、y 軸

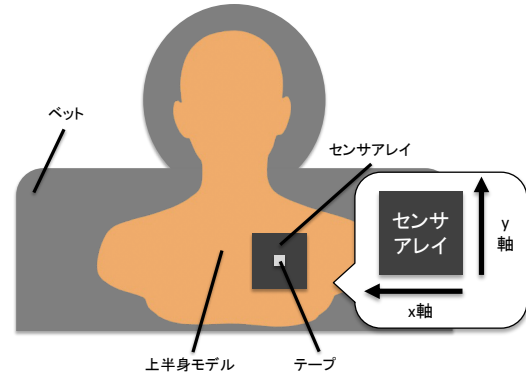


図 1: 実験環境

角度 11.74 であった。

手技療法動作指示語から、対象者が想起した指圧動作における力の強さと角度の時系列データを取得するため、専門家 3 名（表 2 参照）を対象に実験を行った。実験では参加者に 3.1 節で述べた 15 種類の手技療法動作指示語に基づき、各々に対応する指圧動作をセンサアレイに対して実施するよう指示し、センサデータを取得した。取得したデータから、LSTM モデルにより手技療法動作時の圧力と角度（x 軸角度、y 軸角度）の時系列データを推定した。図 1 に実験環境を示す。マッサージ練習用のソフトビニール製の上半身モデルが、マッサージ用のベッド上にうつ伏せに配置されている。また、上半身モデルの上部にセンサアレイが配置されており、参加者はあらかじめセンサアレイ内のテープで指定した部位を指圧動作で押す。手技療法動作指示語をランダムに提示し、各指示から参加者が想起した 15 種類の指圧動作を 1 試行とし、計 3 試行実施した。手技療法動作指示語の提示は、手技療法指導を想定し口頭で行い、試行間には 3 分の休憩を挟んだ。「面に向かって 15 度くらいの角度で押す」という手技療法動作指示語に関しては、分析を容易にするため、x 軸角度に対して 15 度の指圧を行うよう参加者の立つ位置を指定した。各試行における指圧動作の流れを以下に示す。

- (1) 手技療法動作指示語を提示（15 種類各 3 試行）
- (2) 指圧動作の実施
- (3) 動作終了・インターバル（15 秒間）

先述したとおり、手技療法動作指示は口頭で指示され、参加者は自らの主観により指圧動作を行った。

4 分析項目

手技療法動作指示語から、対象者が想起した指圧動作に対する特徴（力の強さ、力の角度、圧力変化、時間）

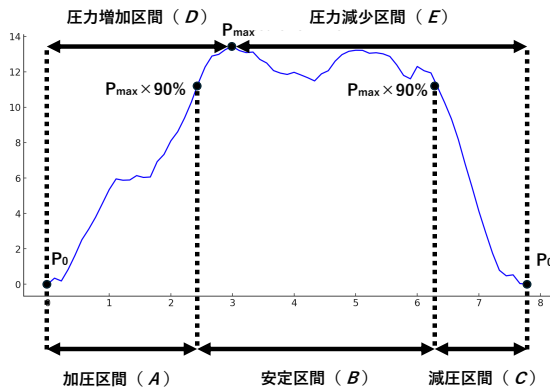


図 2: 区間の定義, 横軸: 時間 (s), 縦軸: 圧力 (N)

の定量的分析を行うため, 実験で取得した時系列データから分析項目を 13 個定義し, これらを算出して分析を行う。

分析項目の定義にあたり, 山下らの研究 [7] を参考に指圧区間を 3 つの区間に分割した。指とセンサが接触している, 0N を超えた地点から再度 0N に戻るまでの区間を指圧区間 (*total*), 0N を超えた地点から 1 回の指圧内で最大圧力値の 90% ($P_{\max} \times 0.9$) に到達するまでの区間を加圧区間 (A), 1 回の指圧内で最大圧力値の 90% ($P_{\max} \times 0.9$) に安定している区間を安定区間 (B), 1 回の指圧内で最大圧力値の 90% ($P_{\max} \times 0.9$) を下回った地点から 0N までの区間を減圧区間 (C) とした。さらに, 1 回の指圧における最大圧力値を参考に指圧区間を 2 つの区間に分割した (図 2 参照)。0N を超えた地点から 1 回の指圧内で最大圧力値 ($P_{\max}$) に到達するまでの区間を圧力増加区間 (D), 1 回の指圧内で最大圧力値 ($P_{\max}$) を下回った地点から 0N までの区間を圧力減少区間 (E) とした。

時間に関する分析項目として, 6 つを定義した。指圧区間 (*total*) の時間を指圧時間, 加圧区間 (A) の時間を加圧時間 (T_A), 安定区間 (B) の時間を安定時間 (T_B), 減圧区間 (C) の時間を減圧時間 (T_C), 圧力増加区間 (D) の時間を圧力増加時間 (T_D), 圧力減少区間 (E) の時間を圧力減少時間 (T_E) とした。

力の強さに関する分析項目として, 海老名らの研究 [3] を参考に安定区間 (B) 内での圧力の平均値を指圧力 (P) と定義した。

力の角度に関する分析項目として, 指圧区間 (*total*) における x 軸角度の平均値を x 軸角度, 指圧区間 (*total*) における y 軸角度の平均値を y 軸角度と定義した。

圧力変化に関する分析項目として, 加圧区間 (A) における圧力の平均変化速度を加圧速度 (v_{press}), 減圧区間 (C) における圧力の平均変化速度を減圧速度 (v_{release}), 加圧区間 (A) 内の圧力速度 ($\frac{dP}{dt}$) の変化率を示す加速度

の最大値と最小値の差を加圧区間 (A) 区間における加速度差 ($A_{\Delta, A}$), 減圧区間 (C) 内の圧力速度 ($\frac{dP}{dt}$) の変化率を示す加速度の最大値と最小値の差を減圧区間 (C) における加速度差 ($A_{\Delta, C}$) と定義した。圧力変化に関する分析項目として, 速度と加速度差を各々 2 つずつ採用した。加圧速度と減圧速度は動作全体におけるマクロな圧力変化を表す分析項目として活用した。加速度差は動作中の急激な圧力変化や滑らかな圧力変化を反映し, 圧力変化率の均一性を表す分析項目として活用した。

5 分析

実験で取得した時系列データに対して, 海老名らの研究 [3] を参考に Savitzky-Golay filter [8] による平滑化 (次数 3, 窓数 15) を行った。その後, 指とセンサが接触していない区間において, 圧力値 0 が連続している部分のデータを接触直前と接触直後の各々の 1 タイムステップのデータ以外を削除した。これらの処理を行った時系列データから 4 節で示した 15 種類の分析項目を算出した。

力の強さに関する 6 種類の手技療法動作指示語と指圧力 (P), x 軸角度に関する 4 種類の手技療法動作指示語と x 軸角度 ($\bar{\theta}_x$), y 軸角度に関する 2 種類の手技療法動作指示語と y 軸角度 ($\bar{\theta}_y$), 圧力変化に関する 5 種類の手技療法動作指示語と圧力変化に関する 4 種類の分析項目について, 各々の関係性の分析を行った。各関係性において, 指示語間における特徴量の差異を検証するため, Friedman 検定を 6 群間で実施することで手技療法指示語が指圧力に与える全体的な影響を評価した (y 軸角度の関係性は 2 種類の動作指示語であるため除く)。有意差が認められた場合には, Wilcoxon の符号付順位検定で多重比較を行い, どの指示語間で差があるかを明らかにした。多重比較の際, Bonferroni 補正を適用し有意水準の補正を行うことで, 第 1 種類の過誤 (Type I エラー) を抑制した。また, 各指示語における各特徴量の個人差を定量的に評価するために被験者ごとに各手技療法動作指示語に対する特徴量の個人内平均値を算出した後, 個人内平均の標準偏差 (以下, 個人内平均標準偏差と記す) を算出した。また, 定量的指示の数値と実際の数値の誤差を定量的に評価するため, 「体重 3kg で押して」に関しては, 理想的な指圧力を 3kg (29.4N), 「面に向かって 15 度くらいの角度で押す」に関しては, 理想的な x 軸角度 ($\bar{\theta}_x$) を 15 度, 「真っ直ぐ上から押す」は理想的な x 軸角度 ($\bar{\theta}_x$) と y 軸角度 ($\bar{\theta}_y$) を 0 度とし, 試行結果と理想値との差を用いて平均二乗誤差 (MSE) を算出した。

表 3: 力の強さに関する表現に対する指圧力の平均値

A	B	C	D	E	F
38.07	11.51	11.12	12.07	28.17	16.70

表 4: 指圧力における力の強さに関する表現間で実施した検定の結果

F 値	p 値	多重比較
30.71	*	A > E > F,D,B,C; F > C

*: $p < 0.05$, 多重比較において, A > B の場合は A と B の間には有意差がある ($p < 0.05$).

次に時間に関する 2 種類の手技療法動作指示語の分析を行った。「3~5 秒かけてじっくり押す」の表現のうち「3~5 秒」という秒数に関する表現は, 安定時間 (T_B), 加圧時間 (T_A) と安定時間 (T_B) の合計時間, 指圧時間 (T_{total}) の 3 つのいずれかを表現していることが考えられる. そのため, 「3~5 秒かけてじっくり押す」の指示語の分析項目を安定時間 (T_B), 加圧時間 (T_A) と安定時間 (T_B) の合計時間, 指圧時間 (T_{total}) を「3~5 秒かけてじっくり押す」の 3 つを選定した. 「3 秒かけてゆっくり押し, 3 秒かけて指の力を抜きながら戻す」の表現は, 加圧時間 (T_A) 3 秒, 減圧時間 (T_C) 3 秒という解釈と, 圧力増加時間 (T_D) が 3 秒, 圧力減少時間 (T_E) P_{max} が 3 秒という解釈が考えられる. まず, 各指示語における各特徴量の個人内平均標準偏差を算出した. 次に, 「3~5 秒かけてじっくり押す」の安定時間 (T_B), 加圧時間 (T_A) と安定時間 (T_B) の合計時間, 指圧時間 (T_{total}) の理想値を各々 4 秒, 「3 秒かけてゆっくり押し, 3 秒かけて指の力を抜きながら戻す」の理想的な加圧時間 (T_A) と減圧時間 (T_C), 圧力増加時間 (T_D), 圧力減少時間 (T_E) を各々 3 秒とし, 試行結果と理想値との差を用いて平均二乗誤差 (MSE) を算出した.

6 結果・考察

6.1 力の強さに関する分析

表 3 に力の強さに関する指示語 (A: 強めに押す, B: 弱めに押す, C: 軽い刺激で押す, D: 少ない刺激で押す, E: 体重 3kg くらいかけて押す, F: 柔らかく押す) に対する指圧力 (P) の平均値を示す. 指圧力の平均値は, A, E, F, D, B, C の順に高い結果となった. 6 種類の指示語間における差異を検証するため, Friedman 検定を 6 群間で実施した結果, 有意な差が認められた ($p < 0.05$). 有意差が認められたため, Wilcoxon の符

表 5: 指圧力の個人内平均標準偏差

A	B	C	D	E	F
10.22	5.26	1.53	6.23	1.33	4.12

表 6: x 軸角度に関する表現に対する x 軸角度 ($\bar{\theta}_x$) の平均値

A	B	C	D
-12.21	12.10	-2.52	9.21

号付順位和検定を用いて多重比較を行った (表 4 参照). 多重比較の結果から, A といった強い指圧を指示する表現, B といった中程度の指圧を指示する表現, C, D, E, F といった弱い指圧を指示する表現を各々区別して, 力加減を行っている可能性が示された. C, D, E, F の指示語は要求される強さが類似しているため, 施術者がこれらの表現を区別できず, 同様の圧力で指圧している可能性が考えられる. 力の強さに関する指示語は漠然とした指圧力は伝達できるものの, 詳細な指圧力の伝達が難しいと考えられる. 各指示語における指圧力の個人内平均標準偏差を表 5 に示す. いずれの指示語下でも一定のばらつきが見られ, 個人差や個人内差が見られた. A の個人内平均の標準偏差が最も高いことから, 指示する指圧が強い場合, 個人差が大きくなる可能性が示された. 一方で, B といった中程度の指圧を指示する表現は, 個人内平均標準偏差が最も小さく, 定量的表現を用いることで個人差が小さくなる可能性が示唆された. 弱い指圧を指示する表現のうち, C, D, E, F の順で個人内平均の標準偏差が小さいことから, 弱い指圧を学習者に指導する場合, C という表現を用いることで誤差が小さくなる可能性が示唆された. 「体重 3kg で押して」という理想値が明確な定量的表現に関しては, 各試行結果と理想値との差を用いて平均二乗誤差 (MSE) を算出した結果, 13.68 となり, 理想値と一定の誤差があることが明らかとなった.

6.2 力の角度に関する分析

表 6 に, x 軸角度に関する手技療法動作指示語 (A: 体の中心方向に向かって押す, B: 外側方向に向かって押す, C: 真っ直ぐ上から押す, D: 面に向かって 15 度くらいの角度で押す) に対する x 軸角度 ($\bar{\theta}_x$) の平均値を示す. x 軸角度 ($\bar{\theta}_x$) の平均値は, B, D, C, A の順に大きい値を示した. 4 種類の指示語間における差異を検証するため, Friedman 検定を 4 群間で実施した結果, 有意な差が認められた ($p < 0.05$). 有意差が認められ

表 7: x 軸角度 ($\bar{\theta}_x$) における x 軸角度に関する表現間で実施した検定の結果

F 値	p 値	多重比較
25.93	*	B > D > C > A

*: $p < 0.05$, 多重比較の結果, A > B の場合は A と B の間に有意な差がある ($p < 0.05$).

表 8: x 軸角度に関する表現に対する x 軸角度 ($\bar{\theta}_x$) の個人内平均標準偏差

A	B	C	D
2.90	0.81	2.08	1.52

表 9: y 軸角度に関する表現に対する y 軸角度 ($\bar{\theta}_y$) の平均値

	A	B
y 軸角度 ($\bar{\theta}_y$)	-0.38	-10.73

たため、Wilcoxon の符号付順位和検定を用いて多重比較を行った (表 7 参照)。多重比較の結果, A, B, C, D の全ての表現を区別して, 動作を行っている可能性が示唆された。各指示語における x 軸角度 ($\bar{\theta}_x$) の個人内平均標準偏差を表 8 に示す。「真っ直ぐ上から押す」は他の表現に比べ理解が容易であると考えられるため, 個人内平均の標準偏差が最も小さい値を示す結果になった。「面に向かって 15 度くらいの角度で押す」は, 個人内平均標準偏差が最も小さく, 定量的表現を用いることで個人差が小さくなる可能性が示唆された。「真っ直ぐ上から押す」「面に向かって 15 度くらいの角度で押す」に関しては, 試行結果と理想的な x 軸角度 ($\bar{\theta}_x$) との差を用いて平均二乗誤差 (MSE) を各々算出した。その結果, 「真っ直ぐ上から押す」の MSE は 12.12, 「面に向かって 15 度くらいの角度で押す」の MSE は 40.17 となり, 理想値と一定の誤差があることが明らかになった。

表 9 に y 軸角度に関する手技療法動作指示語 (A: 真っ直ぐ上から押す, B: 頭方向に向かって押す) に対する y 軸角度 ($\bar{\theta}_y$) の平均値を示す。y 軸角度 ($\bar{\theta}_y$) の平均値は, A, B の順に大きい値を示した。2 種類の指示語間と y 軸角度 ($\bar{\theta}_y$) の差異を検証するため, Wilcoxon の符号付順位和検定を実施した結果, 有意な差が認められた ($p < 0.05$)。A, B の表現を区別して, 動作を行っていること可能性が示唆された。各指示語における y 軸角度 ($\bar{\theta}_y$) の個人内平均標準偏差を表 10 に示す。A は B に比べ理解が容易であると考えられるため, 個人内平均の標準偏差が最も小さい値を示す結果になった。「真っ

表 10: y 軸角度に関する表現に対する y 軸角度 ($\bar{\theta}_y$) の個人内平均標準偏差

	A	B
個人内平均標準偏差	0.54	3.81

表 11: 圧力変化に関する特徴量の平均値

	A	B	C	D	E
v_{press}	8.44	7.17	6.08	6.65	7.47
$v_{release}$	-6.57	-6.63	-9.08	-11.87	-8.37
A_{Δ, T_A}	107.73	89.55	133.67	96.78	89.99
A_{Δ, T_C}	136.20	142.15	288.94	140.41	97.63

表 12: 圧力変化に関する特徴量の個人内平均の標準偏差

	A	B	C	D	E
v_{press}	3.89	2.81	3.84	1.02	2.81
$v_{release}$	2.89	0.57	4.76	7.49	2.57
A_{Δ, T_A}	33.76	44.08	64.14	25.49	31.80
A_{Δ, T_C}	70.78	171.18	308.98	41.25	30.50

直ぐ上から押す」に関しては試行結果と理想値との差を用いて平均二乗誤差 (MSE) を算出した結果, 4.96 となり, 理想値と一定の誤差があることが明らかとなった。

6.3 圧力変化に関する分析

表 11 に, 圧力変化に関する手技療法動作指示語 (A: 3 秒～5 秒かけてじっくり押す, B: 3 秒かけてゆっくり押し, 3 秒かけて指の力を抜きながら戻す, C: じんわり圧をかけて押す, D: グーッと押す, E: 柔らかく押す) に対する圧力変化に関する 4 つの特徴量の平均値を各々算出した結果を示す。また, 表 12 に個人内平均標準偏差を算出した結果を示す。各特徴量において, いずれの指示語下で一定のばらつきが見られ, 個人差が見られた。5 種類の指示語間における圧力変化に関する 4 つの特徴量の差異を各々検証するため, Friedman 検定を 5 群間で実施した結果, 減圧速度 ($v_{release}$) のみに有意差が認められた。有意差が認められたため, Wilcoxon の符号付順位和検定を用いて多重比較を行った。D と B のペアのみ有意差が認められ ($p < 0.05$), 他のペアでは有意差が認められなかった。

6.4 時間に関する分析

表 13 に「3～5 秒かけてじっくり押す」に対する各特徴量 (安定時間 (T_B), 加圧時間 (T_A) と安定時間 (T_B) の合計時間, 指圧時間 (T_{total})) の平均値, 個人内平均

表 13: 「3 秒～5 秒かけてじっくり押す」における各特徴量 (T_B , $T_A + T_B$, T_{total}) の分析結果

	平均値	個人内平均標準偏差
T_B	0.94	0.24
$T_A + T_B$	3.17	0.50
T_{total}	6.12	1.23

表 14: 「3 秒～5 秒かけてじっくり押す」における被験者ごとの各特徴量 (T_B , $T_A + T_B$, T_{total}) の平均二乗誤差 (MSE)

	実験参加者 1	実験参加者 2	実験参加者 3
T_B	8.82	11.14	9.23
$T_A + T_B$	1.95	0.21	1.21
T_{total}	0.75	8.94	7.74

標準偏差を示す。いずれの指示語下でも一定のばらつきが見られ、個人差が見られた。表 14 に被験者ごとの、各特徴量の理想値との平均二乗誤差を示す。この結果から、実験参加者 1 は「3～5 秒」という秒数に関する表現は指圧時間 (T_{total}) を表現していると認識している可能性が高く、最も平均二乗誤差が小さい値を示す結果となった。実験参加者 2 と実験参加者 3 は「3～5 秒」という秒数に関する表現は加圧時間 (T_A) と安定時間 (T_B) 合計時間を表現していると認識されており、最も平均二乗誤差が小さい値を示す結果となり、約 4 秒指圧を行った後に力を抜いている可能性が示唆された。

表 15 に「3 秒かけてゆっくり押し、3 秒かけて指の力を抜きながら戻す」に対する各特徴量 (加圧時間 (T_A), 減圧時間 (T_C), 圧力増加時間 (T_D), 圧力減少時間 (T_E)) の平均値、個人内平均標準偏差を示す。いずれの指示語下でも一定のばらつきが見られ、個人差が見られた。表 16 に被験者ごとの各特徴量の理想値との平均二乗誤差を示す。この結果から、実験参加者 2 と実験参加者 3 は「3 秒かけてゆっくり押し、3 秒かけて指の力を抜きながら戻す」という表現を圧力増加時間 (T_D), 圧力減少時間 (T_E) が 3 秒であると認識している可能性が示唆された。

7 議論

7.1 手技療法教育のあり方

6 章で行った分析結果により、伝達したい動作に対して効果的な手技療法動作指示語を選定できる。手技療法動作指示語の選定基準は大きく、(1) どのような手技療法動作 (e.g., 力の強さ, 力の角度, 圧力変化, 指圧時間)

表 15: 「3 秒かけてゆっくり押し、3 秒かけて指の力を抜きながら戻す」における各特徴量 (T_A , T_B , T_C , T_D) の分析結果

	平均値	個人内平均標準偏差
T_A	3.17	0.68
T_B	2.96	0.34
T_C	3.38	0.59
T_D	3.14	0.34

表 16: 「3 秒かけてゆっくり押し、3 秒かけて指の力を抜きながら戻す」における被験者ごとの各特徴量 (T_A , T_B , T_C , T_D) の平均二乗誤差 (MSE)

	実験参加者 1	実験参加者 2	実験参加者 3
T_A	0.16	0.33	4.06
T_C	0.17	0.50	2.05
T_D	0.44	0.27	3.74
T_E	0.05	0.07	1.90

を学習者に行ってほしいか、(2) 動作の個人差が小さい表現を選定する、の 2 つに分けられる。具体的には、(1) 指導者が弱い指圧を学習者に指示する際に、「柔らかく押す」「少ない刺激で押す」「弱めに押す」「軽い刺激で押す」が選定される。次に、(2) 「軽い刺激で押す」「柔らかく押す」「弱めに押す」「少ない刺激で押す」の順で個人差が小さいため、「軽い刺激で押す」という表現を選定する。これにより、弱い指圧動作を指導する際、学習者と指導者の意図する動作の認識の差異を小さくすることが可能である。

次に、従来の手技療法動作の指導方法と、手技療法動作定量化システムを活用した指導方法の利点と欠点を整理する。動作指示語による指導方法では、正確に伝達できない情報 (e.g., 精密な力の強さ, 精密な圧力変化) が存在しているため、学習者にとって理解が難しい場合も少なくない。さらに、学習者が指導者の意図した動作を正確に理解した場合であっても、その動作が再現できているか確認することが難しい。一方で、動作指示語での指導は漠然とした動作のイメージ (e.g., 「早く押して」の場合、直観的に素早い動きをイメージする) を直観的に理解することが可能である。また、機器を設置する必要がなく、伝達方法が簡便である。手技療法動作定量化システムを活用した指導方法では、こうした直観的で簡便な伝達は難しい。一方で、手技療法定量化システムでは従来の指導方法では正確に伝達できない動作の精密な力の強さ, 精密な圧力変化等の情報を理解することができる。また、海老名ら [3] によると主観的な指示 (弱く、

普通の力で、強くの3種類の指示)では指圧力は個人差がある一方で、指圧力を定量化するシステムを用いて、定量的指示を行った際、高い精度で目標値を再現できることが報告されている。さらに、これらの数値を参考にすることで、学習者が指導者の意図した動作を正確に再現できているか確認できる[4]。これらの各指導方法の利点と欠点を踏まえ、学習者の特性等を考慮し指導方法を活用していくことで、手技療法指導の質向上に繋がると考えられる。具体的には、手技療法の初学者に対しては、従来の指導方法で主観的な表現で漠然とした動作のイメージをつかませることで基礎技能を習得させ、熟達者に対しては、定量化システムを活用し精密な技術を習得させることが考えられる。

7.2 今後の展望

今後は各特徴量に関する表現を増やし、主観的な表現ではどの程度の粒度で手技療法動作(e.g., 力の強さに関する指示語は「弱めの指圧力」「中程度の指圧力」「強めの指圧力」の3つの指圧力を伝達可能である)を伝達できるか検証を行う。実験対象者は3名は全員が手技療法の経験が2年以上であった。そのため、手技療法初心者を対象とした際の分析を行い比較することで、熟達者と初心者の手技療法の差異の分析が可能になる。また、分析結果より、「体重3kgくらいかけて押す」「面に向かって15度くらいの角度で押す」は、定量的表現を用いることで個人差が小さくなる可能性が示唆されたが、定量的表現の度合いによって、同じ結果が得られるか明らかになっていない。今後は、定量的表現を増やし、異なる度合いの表現においても個人差が小さくなるか確認することを検討する。また、他の手技療法動作として、なでる、もむ、叩く等の手技療法動作を対象とし、これらの動作に関連する手技療法動作指示語から、対象者が想起した動作を分析することを検討する。

8 おわりに

本稿では、手技療法における動作と動作指示語の対応関係を明らかにした。専門家3名を対象とした実験では、それぞれの強さや角度等を示す動作指示語は使い分けて表現されており、システムによる計測値でも一定の有意差が確認された。これにより、これまで感覚的な言語表現に依存していた圧力・角度を数値情報で補完できる可能性が示唆された。今後は、対象動作指示語、対象動作、対象参加者を変更し、分析することを試みる。

謝辞

本研究の一部は、関西大学医工薬連携研究費の支援を受けた。記して謝意を表す。

参考文献

- [1] 矢野忠, 安野富美子, 藤井亮輔, 鍋田智之: あん摩マッサージ指圧療法, 鍼灸療法に対する受療者の評価に関する調査(後編), 医道の日本, Vol. 79, No. 7, pp. 180-187 (2020).
- [2] 篠原和也, 葛谷憲彦, 鈴木ゆい, 鹿田将隆: 理学療法士, 作業療法士, 言語聴覚士の臨床能力および技能を測定する評価ツールに関するナラティブレビュー, 日本保健科学学会誌, Vol. 26, No. 3, pp. 150-160 (2023).
- [3] 海老名光希, 後藤充裕, 瀬下仁志: 遠隔触診に向けた程度副詞指示に対する指圧力個人差および定量的指示下での指圧力再現性に関する検証, 電子情報通信学会 HCG シンポジウム 2024 論文集, pp. C-4-1 (2024).
- [4] 田中瑠慧, 松下光範: 接触抵抗式感圧センサを用いた手技療法動作の定量化, 情報処理学会研究報告, Vol.2025-HCI-211, No.44, pp.1-8, (2025).
- [5] 北尾浩和, 来田宣幸, 深田智, 中本隆幸, 小島隆次, 萩原広道, 野村照夫: 言語的な動作指示の違いがパフォーマンスに及ぼす影響, 日本感性工学会論文誌, pp. TJSKE-D-17-00039(2017).
- [6] 金田有美子, 葛山元基, 小林武雅, 古谷美帆, 吉田大記, 成命奇, 勝平純司: 平行棒を使用した立ち上がり動作時の関節モーメントの分析, 理学療法科学, Vol. 21, No. 3, pp.227-232 (2006).
- [7] 山下達也, 中尾恵, 松田哲也: つまみ及びつかみ操作における指先の動的特徴量に関する定量分析, 生体医工学, Vol. 57, No. 2-3, pp. 68-74 (2019).
- [8] Savitzky, A. and Golay, M. J.: Smoothing and differentiation of data by simplified least squares procedures., Analytical chemistry, Vol. 36, No. 8, pp. 1627-1639 (1964).

LLMによる説明文の生成と対照学習を用いた 日本酒-料理の相互推薦

Mutual Recommendation of Sake and Foods Using LLM-Based Descriptive Text Generation and Contrastive Learning

松田颯太^{1*} 尾崎 知伸¹
Souta Matsuda¹ Tomonobu Ozaki¹

¹ 日本大学文理学部情報科学科

¹ Department of Information Science, Nihon University

Abstract: This study proposes a mutual sake-food recommendation system to support consumers in judging pairings. The proposed system (1) generates sentences using an LLM for sake and food information and pairing suggestions, (2) embeds these sentences into a vector space via document embedding and contrastive learning based on sake-food relationships, and (3) finally recommends suitable sake-food pairs within this space. The proposed system, using approximately 1000 of sake and 600 of food, was evaluated both quantitatively through computational experiments and qualitatively through a questionnaire survey. The results showed that the proposed system has a certain degree of effectiveness.

1 はじめに

近年、観光や外食産業、個人の食習慣において、飲食物の組み合わせに対する理解の重要性が増している。日本酒と料理のペアリングは、日本の食文化において重要な要素であり、適切な組み合わせが双方の味わいを最大化するとされる。しかし日本酒は一般に、地域や製造方法によって味わいや香りが大きく異なるため、一般消費者が料理との相性を判断することは容易ではない[1]。また、日本酒離れの傾向も顕著であり、ある調査では若年層の70%、女性の74%が直近1年間に日本酒を飲んでいないことが報告されている[2]。本研究では、このような状況を改善し、また日本酒への関心を高めると共に料理との組み合わせを含めた具体的な楽しみ方を提案するための基礎技術として、日本酒と料理の特性を考慮した日本酒-料理の相互推薦システムを構築することを目的とする。

日本酒は、精米歩合、日本酒度、酸度、アルコール度数などの化学的特性に加え、風味や香りといった感覚的特性を持つ。一方で料理は、調理方法、味覚、香り、食材の組み合わせといった多様な要素・視点を含む。精度の高い相互推薦システムを実現するためには、個々の日本酒や料理が持つ詳細な特徴や多様な観点を考慮する必要がある。詳細は後述するが、本研究では、既

存手法である LLM-Rec[3] のアイディアに基づき、種々の観点における日本酒・料理の詳細情報を推薦に反映させることを試みる。具体的には、大規模言語モデル (Large Language Model, LLM) を用いて各日本酒・料理の情報を補完・拡張した説明文を生成した上で、文書埋め込み技術を用いて日本酒・料理をベクトル空間へと埋め込む。加えて、相性を考慮した日本酒-料理対や、類似性を考慮した日本酒-日本酒対と料理-料理対を学習データとして準備し、埋め込み空間上で対照学習を展開することで、日本酒・料理それぞれに関する埋め込みモデルを学習する。得られた埋め込みモデルを利用し、ベクトル空間上で類似度の高い日本酒-料理対を推薦対象とする。これらを通じ、日本酒・料理の詳細情報と多様な視点を考慮した、精度の高い相互推薦の実現を目指す。

本論文の構成は以下のとおりである。2章で関連研究について概説する。3章で提案手法を述べる。4章で評価実験について報告し、最後に5章でまとめと今後の展望を述べる。

2 関連研究

これまでに LLM を活用した推薦システムが数多く提案されている。

推薦において、対象を表現するタグ情報の設定は重要

*連絡先：日本大学文理学部情報科学科
〒156-8550 東京都世田谷区桜上水 3-25-40
E-mail: chso21022@g.nihon-u.ac.jp

な要素となる。このことに関連し、文献 [4] では、LLM を活用することによりマルチモーダルデータからタグを生成・抽出するシステム TagGPT が提案されている。また TagGPT はプロンプトを通じ、ユーザーの意図や関心を反映した高品質なタグ集合の自動構築を実現している。

一方、テキストに基づく推薦システムでは、推薦アイテムに関する記述の不完全さが、推薦精度の劣化を招く大きな要因となる。この問題を軽減するため、文献 [3] では、LLM を用いた説明テキストの拡張・生成を伴う推薦システム LLM-Rec が提案されている。LLM-Rec は、LLM に対する多様なプロンプトを用いて種々の側面からアイテムの記述情報を補完するとともに、これらの情報とユーザープロフィールとを擦り合わせることで、推薦の個人化を実現している。また文献 [5] では、利用者が自然言語を用いて自由にニーズを表現できる推薦システムの実現を目指し、LLM による指示遂行として推薦を実現する枠組み InstructRec を提案している。具体的には、柔軟で汎用的な指示形式を設計した上で、LLM を推薦システムに特化した形でファインチューニングすることで、順序推薦や個人化検索といった多様な推薦タスクに対応している。

LLM を活用する推薦システムでは、プロンプトの質が推薦システムの質へと直結する。プロンプト最適化手法として、文献 [6] では、テキスト形式で表現されるプロンプトをベクトル系列として表現される連続プロンプトへと変換するプロンプト蒸留 (PrOmpt Distillation, POD) アプローチが提案されている。これにより、LLM の汎用性を維持しつつ、ユーザーの意図やコンテキストをより適切に捉えた高品質な推薦の実現が期待される。

以上概観した通り、LLM の活用した推薦システムの高度化技術は多岐にわたる。本研究では、これらの高度化技術の中でも特に LLM-Rec[3] の考えに焦点を当て、推薦対象となる日本酒・料理の情報を拡張・補完した上で、ベクトル空間へと埋め込むアプローチを採用する。

3 提案手法

3.1 概要

本研究では、以下に示す 3 つの手順に従い、詳細な特徴や種々の観点を反映した日本酒と料理に関する埋め込みモデルを構築することで、同一埋め込み空間上での距離に基づく相互推薦システムを実現する。図 1 に、提案システムの全体像を示す。なお LLM には OpenAI の GPT-4o[7] を、文書埋め込みモデルには Sentence-Bert[8] をそれぞれ用いる。

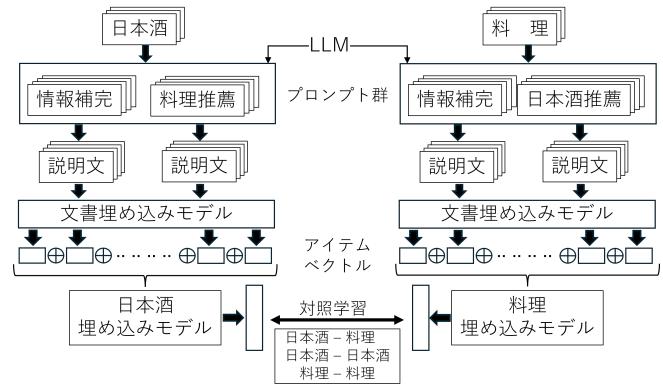


図 1: 提案手法の概要

1. 複数の観点からそれぞれの情報を引き出すプロンプトを LLM に与えることで、各アイテム（日本酒・料理）の詳細を補完・拡張する説明文集合を生成する。
2. 生成した各観点からの説明文を、文書埋め込みモデルを用いてそれぞれベクトル化する。また、これらのベクトルをアイテム毎に連結することで、アイテムベクトルを構築する。
3. 学習データとして、相性の良し悪しをクラスとする日本酒-料理対、および類似性の有無をクラスとする日本酒-日本酒対と料理-料理対を準備し、アイテムベクトルに関する対照学習 [9] を行うことで、日本酒と料理に対する各埋め込みモデルを構築する

埋め込みモデル構築において、アイテムベクトルの構築（手順 1 と手順 2）は、本質的に LLM-Rec[3] における方法を援用するものではあるが、日本酒や料理に特化した説明文生成プロンプトを準備することで、精度の向上を試みる。また対照学習（手順 3）では、日本酒や料理に関する知識を基とした学習データの準備が必要となる。それぞれについて、以下の節で詳細を説明する。

3.2 LLM を用いた説明文の生成

本研究では、日本酒と料理の特性を多角的に分析し、最適なペアリングを提案するため、日本酒の特性に基づく料理推薦と、料理の特性に基づく日本酒推薦の 2 つの視点からプロンプトを準備した。さらに、各アイテム（日本酒・料理）の情報補完を目的としたプロンプトを組み合わせることで、提案の精度を高めることを目指した。以下、それぞれについて説明する。

3.2.1 日本酒に関するプロンプト

日本酒に関するプロンプトとして、日本酒情報補完プロンプトと料理推薦プロンプトの2種類を準備する。

日本酒情報補完プロンプト： このプロンプトは、日本酒の特性を深く理解し、料理との相性を科学的に説明するために準備した。料理との相性を説明させることで、ペアリング提案の精度を向上させることが期待される。具体的には、LLMに、多角的な日本酒特性の解釈と、特性間の相互作用を分析させることにより、精米歩合や酸度が風味に与える影響や日本酒度が料理とのバランスに与える効果、飲用温度による味覚の変化を考察し、それぞれの観点からペアリング提案の基礎情報を提供させる4つのプロンプトから構成される。

料理推薦プロンプト： 日本酒の特性は料理との相性に大きく影響を与える。そこで、LLMに対して数値的特性を分析させることで、科学的根拠のある具体的な提案をさせることを試みる。また、提案の多様性と実用性の確保を目的に、飲用温度や日本酒タイプによる味覚の変化を考慮させることとした。具体的には日本酒の特性（精米歩合、日本酒度、酸度、アルコール度数など）や、日本酒タイプ（薫酒、爽酒、醇酒、熟酒）、飲用温度に基づき、例えば、酸味や旨味を強調する料理や、特定の調理法に適した料理を提案させるなど、各特性を引き立てる料理を提案する4つのプロンプトから構成される。

3.2.2 料理に関するプロンプト

日本酒に関するプロンプトと同様、料理に関するプロンプトとして、情報補完プロンプトと日本酒薦プロンプトの2種類を準備する。

料理情報補完用プロンプト： 4つのプロンプトから構成され、LLMに対し、料理に使用される材料や調理法、味覚の詳細を補完させると共に、各味覚が料理全体に与える影響や特定の食材・調理法が風味に与える効果、スパイスやハーブの香りが料理に与える役割を分析させ、日本酒との調和を考察させる。これにより、料理の特性を深く理解し、日本酒との相性をより正確に評価するため情報を提供が期待される。

日本酒推薦プロンプト： 料理の味覚や香りが日本酒との相性に強く影響を与えること考慮し、LLMに対し、調理法や食材の特性が日本酒の特性とどのように調和するかを分析させた上で、当該外料理に合う日本酒を推薦させる説明文を生成させる。具体的には、料理の基本味覚（甘味、酸味、苦味、塩味、旨味）、主要食

材、調理方法、香りに基づき、甘味や酸味の強い料理にはそれを調和させる日本酒を、スパイスの効いた料理には香りを補完する日本酒を選定など、料理を引き立てる日本酒を提案させる4つのプロンプトから構成される。

3.3 対照学習を通じた埋込モデルの構築

本研究では、日本酒・料理の各埋込込みモデルとして、文書埋込込みモデル（Sentence-Bert）の出力を連結したベクトルを受け取り、それを64次元の潜在空間へとマッピングするニューラルネットワークを構築・学習する。

各モデルは、入力層に加え、256次元の全結合中間層と64次元の全結合出力層の3層構造を基本とし、ReLU活性化関数、バッチ正規化、Dropoutを組み合わせることで、計算効率を維持しつつ、学習の安定性と高い汎化性能の両立を目指す構成を採用している。また学習には、コントラストロスと損失関数とする対照学習の枠組みを採用している。すなわち、学習データとして与えられる、日本酒-料理対および日本酒-日本酒対、料理-料理対の各対に対し、正例対に含まれるアイテム同士のベクトルは近くに、負例対に含まれるアイテム同士のベクトルは遠くにそれぞれ配置されるよう、各埋込込みモデルの重みを学習する。

4 評価実験

4.1 データセット

本研究では、日本酒に関するデータセットとして、“Sake dataset”[10]を使用した。このデータセットは約1,000件の日本酒情報から構成されており、各日本酒に関する属性として、銘柄、品目、酒造、生産地情報（都道府県、地域）、分類、造り、精米歩合、アルコール度数、日本酒度、酸度、アミノ酸度、甘辛度、使用米、酵母などが含まれている。このデータセットを対象に、欠損値を含むデータを除外し、各酒造サイトに掲載されている各日本酒に対する説明を追加することで、評価実験用のデータセットを構築した。

一方、料理に関するデータセットには、クックパッド株式会社が国立情報学研究所と協力して研究者に提供しているクックパッドデータセット[11]を使用した。各料理データには、料理名に加え、食材と調理手順の情報が含まれている。実験では、データセット全体から「様々な種類の料理を選ぶこと」を基準に620件の料理データを抽出して利用する。

準備した日本酒データおよび料理データを対象に、対照学習用のデータセットを構築する。日本酒-料理対は、

表 1: 対照学習用データのサイズ

	正例対数	負例対数
日本酒 - 料理	2,241	4,119
日本酒 - 日本酒	4,284	3,288
料理 - 料理	1,075	1,329

文献 [12, 13] を参考に相性が良いとされるペアリングを正例対として抽出した。一方、日本酒-日本酒対は、特定名称分類で類似する日本酒を絞り込み、さらにアルコール度数・精米度合いが近いものを選択することで正例対を準備した。また、料理-料理対に関しては、和食・洋食・中華料理などの料理ジャンルと肉料理・魚料理などのメイン食材別の分類を基準に正例対を準備した。なお、いずれも負例対は、正例対に含まれないアイテムの組み合わせをランダムに選択することで準備した。表 1 に、対照学習用データセットに含まれる正負例対数を示す。

4.2 推薦精度

提案手法の性能を評価するため、データセット全体を 7:1:2 の割合で学習例集合: 検証例集合: テスト例集合の 3 つに分け、エポック数を 10 または 20 として各埋め込みモデルの学習を行った。また比較のために、対照学習に利用するデータセットを変えた 4 種のモデル

SF : 日本酒-料理対のみから学習したモデル

SFS : 日本酒-料理対および日本酒-日本酒対から学習したモデル

SFF : 日本酒-料理対および料理-料理対から学習したモデル

SFSF : 日本酒-料理対および日本酒-日本酒対, 料理-料理対から学習したモデル

を構築している。評価尺度には、テスト例集合における平均コサイン類似度、適合率、再現率、NDCG@10 を採用し、再現率・適合率の計算に必要な閾値は検証例集合を基準に決定している。表 2 に実験結果を示す。

表 2 より、提案モデルである SFSF は、適合率において SFS・SFF に及ばないものの、その他の指標において他のモデルを上回る高い性能を示していることが確認できる。また適合率が高い SFF は、再現率に関して性能が低く、特定の関係性への依存が示唆される。これらの結果から、提案モデルが十分な性能を有することに加え、モデルの学習において、複数の関係性を総合的に利用することが有効であることが示された。

表 2: モデル性能の比較

モデル	平均 cos 類似度	適合率	再現率	NDCG @10
エポック数 = 10				
SF	0.2677	0.4379	0.3630	0.3712
SFS	0.3152	0.4511*	0.4128*	0.5266
SFF	0.2823	0.4606*	0.3215*	0.4337
SFSF	0.4654	0.4492*	0.5885**	0.5392
エポック数 = 20				
SF	0.2974	0.4481	0.3643	0.4263
SFS	0.3275	0.4659*	0.4605*	0.5431
SFF	0.2700	0.4935*	0.3259*	0.6058
SFSF	0.4110	0.4561*	0.5265**	0.6459

SF を基準とし *: $p < 0.05$, **: $p < 0.01$

4.3 プロンプト評価

本研究で採用した情報補完および推薦プロンプトの有効性を評価するために、説明文生成に利用するプロンプトを制限した上でモデルの学習を行う比較実験を行った。以下に比較対象として利用するプロンプトを示す。また実験結果、すなわち日本酒-料理対および日本酒-日本酒対, 料理-料理対から学習したモデルの精度を表 3 に示す。

Sr_N : 日本酒に対して、料理推薦プロンプトを用いる。料理に関しては説明文生成を行わず、タイトル, 食材, 手順を連結した文する。

Sc_N : 日本酒に対して、情報補完プロンプトを用いる。料理に関しては説明文生成を行わず、タイトル, 食材, 手順を連結した文する。

SrSc_N : 日本酒に対して、料理推薦プロンプトおよび情報補完プロンプトを用いる。料理に関しては説明文生成を行わず、タイトル, 食材, 手順を連結した文する。

N_Fr : 料理に対して、日本酒推薦プロンプトを用いる。日本酒に関して説明文生成を行わず、説明文をベクトル化する。

N_Fc : 料理に対して、情報補完プロンプトを用いる。日本酒に関しては説明文生成を行わず、説明文をベクトル化する。

N_FrFc : 料理に対して、日本酒推薦プロンプトおよび情報補完プロンプトを用いる。日本酒に関して説明文生成を行わず、説明文をベクトル化する。

Sr_Fr : 日本酒に対して、料理推薦プロンプトを用いる。料理に対して、日本酒推薦プロンプトを用いる。

表 3: 利用プロンプト別の精度

プロンプト	平均 cos 類似度	適合率	再現率	NDCG @10
Sr_N	0.3314	0.3416	0.2959	0.3882
Sc_N	0.3570	0.4478	0.5607	0.5022
SrSc_N	0.3924	0.4453	0.4826	0.3821
N_Fr	0.4054	0.3637	0.5272	0.5646
N_Fc	0.4056	0.3598	0.5189	0.5384
N_FrFc	0.4110	0.3421	0.4265	0.4759
Sr_Fr	0.4116	0.4463	0.5067	0.5452
Sc_Fc	0.4231	0.4754	0.5623	0.5787

Sc_Fc : 日本酒に対して、情報補完プロンプトを用いる。料理に対して、情報補完プロンプトを用いる。

実験結果より、Sc_N や Sc_Fc など、日本酒情報補完プロンプトを含む場合は、平均コサイン類似度・適合率・NDCG@10 が全体的に向上する傾向が確認でき、特に Sc_Fc において、NDCG@10 が 0.5787 を記録し、他の条件を上回った。一方、N_Fr や N_Fc など、料理プロンプトのみを利用した場合は性能が低く、日本酒に関して説明文を補完・生成することが重要であることが示唆された。また、この実験におけるすべての条件において、すべてのプロンプトを利用した場合に対する精度劣化が確認でき、精度の高い相互推薦を行うためには日本酒・料理の双方において説明文を補完・生成することが有効であることが示された。

4.4 定性評価

提案手法を定性的に評価するため、日本酒の関連業界での就業経験を有し、日本酒と料理のペアリングに関して一定の知識を持つ男女 10 名を対象に、以下に示す 2 種類のアンケート調査（設問数はそれぞれ 10）を実施した。

アンケート 1 : 各設問において、1 つの日本酒と 5 つの料理を提示し、提示された日本酒と最も合うと考えられる料理を選択してもらう。

アンケート 2 : 各設問において、日本酒・料理対を提示し、その組み合わせの良さを 5 段階リッカード尺度（1：非常に良い、2：良い、3：普通、4：良くない、5：非常に良くない）で回答してもらう。

両アンケートともに、提案手法で算出される類似度スコアに基づき、不自然さを感じない範囲で比較的上位に含まれる 5 件の日本酒・料理対を上位ペア群、比較的下位に含まれる 5 件を下位ペア群として抽出、利用している。アンケート 1 では、抽出された各上位（下

表 4: アンケート 1 の結果：各日本酒・料理対の得票数

日本酒	料理_1	料理_2	料理_3	料理_4	料理_5
上位ペア					
日本酒_P1	6	2	1	1	0
日本酒_P2	6	3	1	0	0
日本酒_P3	5	3	1	1	0
日本酒_P4	5	2	2	1	0
日本酒_P5	3	4	2	1	0
下位ペア					
日本酒_N1	6	2	0	0	2
日本酒_N2	4	3	0	0	3
日本酒_N3	4	2	1	0	3
日本酒_N4	3	3	1	0	3
日本酒_N5	7	2	1	0	0

位）ペアについて、ペアに含まれる料理を基準とし、相性が良い（悪い）と思われる範囲の中で日本酒との相性の降順（昇順）に料理を 4 件選択することで設問を準備した。

アンケート 1 の結果を表 4 に示す。表中において、下線は設問作成に基となった上位・下位ペアを表す。また、表中の料理 1～料理 5 は、各設問において提示される 5 件の料理に対応し、付与された番号は提案手法により日本酒との相性が良いと判断された順序を表す。実験結果より、上位ペアでは提案手法が提示した日本酒・料理対に対して高い評価が集まっていることが分かる。一方、下位ペアに関しては、料理に対する評価が分散する傾向が認められ、相性が悪いペアリングに対しては専門家でも優劣を判断することが難しいことが伺える。

アンケート 2 では、上位ペア・下位ペアを直接被験者に提示した。結果を表 5 に示す。結果より、上位ペアに関しては、すべての日本酒・料理対において「1：非常に良い」と「2：良い」の和が 50% 以上であることが分かる。また、「5：非常に良くない」との回答はなく、「4：良くない」と合わせても、否定的な意見は各対において最大 30% に抑えられている。一方、下位ペアに関しては、回答が分散してはいるが、「1：非常に良い」と「2：良い」の和は最大で 40% であり、アンケートに回答した半数以上の専門家がポジティブな印象を抱いていないことが伺える。これらの結果から、提案手法を用いることで、人間の直感に概ね合致する日本酒・料理の相互推薦が実現できると考えられる。

5 まとめ

本研究では、大規模言語モデル（LLM）を活用した日本酒と料理の相互推薦手法を提案する。日本酒と料

表 5: アンケート 2 の結果: 回答数

	リッカード尺度の値				
	1	2	3	4	5
上位ペア					
日本酒-料理対_P1	0	5	4	1	0
日本酒-料理対_P2	0	6	2	2	0
日本酒-料理対_P3	0	5	2	3	0
日本酒-料理対_P4	6	2	2	0	0
日本酒-料理対_P5	2	4	4	0	0
下位ペア					
日本酒-料理対_N1	2	1	2	3	2
日本酒-料理対_N2	0	3	2	4	1
日本酒-料理対_N3	2	2	3	3	0
日本酒-料理対_N4	2	1	7	0	0
日本酒-料理対_N5	0	1	5	4	0

理の特性を的確に捉えるため、複数の観点からプロンプトを設計し、各項目の特徴を詳細に記述した文章を生成する。さらに、対照学習 (contrastive learning) の枠組みを活用し、生成された記述文を同一ベクトル空間に埋め込むことで、日本酒と料理の類似性を適切に表現するモデルを構築した。

提案手法の有効性を検証するため、定量的な実験を実施し、推薦精度の向上を確認した。特に、日本酒-料理対のみならず、日本酒-日本酒対や料理-料理対の関係を学習に取り入れることで、埋め込みの精度が向上することを明らかにした。また、プロンプトの設計が推薦精度に大きく影響を与えることを示し、効果的なプロンプト設計の重要性を実証した。

さらに、定性的な評価を通じ、提案手法による推薦結果が人間の直観と整合的であることを確認した。これは、本手法が日本酒と料理の特性を適切に捉え、実用的な推薦を実現できることを示唆している。

今後の課題として、プロンプトのさらなる多様化と外部データの統合が挙げられる。特に、料理の調理法や食文化、地域性などの情報を考慮することで、より精度の高い推薦が可能となると考えられる。また、個人の嗜好や状況に応じた推薦を実現するため、ユーザーのフィードバックを取り入れたパーソナライズド推薦や、文脈依存型の推薦手法の開発を進める予定である。

謝辞: 本研究の遂行にあたり、アンケートにご協力いただいた皆様に心より感謝申し上げます。本研究では、国立情報学研究所の IDR データセット提供サービスによりクックパッド株式会社から提供を受けた「クックパッドデータセット」を利用した。

参考文献

- [1] 石まつ 3 代目: 料理と日本酒のマッチングの会, <http://ishimatsu.cc/archives/857>, 2017. (最終アクセス日: 2025 年 02 月 15 日)
- [2] 楯の川酒造株式会社: 「全国の男女に聞く「日本酒」の飲用実態調査・楯の川酒造調べ」, <https://prtimes.jp/main/html/rd/p/000000030.000071925.html>, 2022. (最終アクセス日: 2025 年 02 月 15 日)
- [3] H. Lyu, S. Jiang, H. Zeng, Y. Xia, Q. Wang, S. Zhang, R. Chen, C. Leung, J. Tang, and J. Luo: LLM-Rec: Personalized Recommendation via Prompting Large Language Models, In *Findings of the Association for Computational Linguistics: NAACL 2024*, pp.583–612, 2024.
- [4] C. Li, Y. Ge, J. Mao, D. Li, and Y. Shan: TagGPT: Large Language Models are Zero-shot Multimodal Taggers, arXiv preprint arXiv:2304.03022, 2023.
- [5] J. Zhang, R. Xie, Y. Hou, X. Zhao, L. Lin, and J.-R. Wen: Recommendation as Instruction Following: A Large Language Model Empowered Recommendation Approach, *ACM Transactions on Information Systems*, 2024.
- [6] L. Li, Y. Zhang, and L. Chen: Prompt Distillation for Efficient LLM-based Recommendation, *Proc. of the 32nd ACM International Conference on Information and Knowledge Management*, pp.1348–1357, 2023.
- [7] OpenAI et, al.: GPT-4 Technical Report, arXiv preprint arXiv:2303.08774, 2023.
- [8] N. Reimers and I. Gurevych: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, *Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp.3982–3992, 2019.
- [9] D. Chicco: Siamese Neural Networks: An Overview, In: Cartwright, H. (eds), *Artificial Neural Networks*, pp.73–94, 2021.
- [10] sake_dataset: Sake (Japanese Rice Wine) dataset / 日本酒データセット. 銘柄ごとにアルコール度数・精米歩合・日本酒度等を数値化した整形済みデータ, https://github.com/yoichi1484/sake_dataset, (最終アクセス日: 2025 年 02 月 15 日)
- [11] クックパッド株式会社: クックパッドデータ, 国立情報学研究所情報学研究データリポジトリ, <https://doi.org/10.32130/idr.5.1>, 2015.
- [12] 葉石かおり: 『日本酒のペアリングがよくわかる本』, シンコーミュージック, 2017.
- [13] 美味しい日本酒: 日本酒のペアリングをタイプ別に解説! 具体例やおすすめ料理も紹介, <https://azumarikishi.co.jp/media/pairing>, 2022. (最終アクセス日: 2025 年 02 月 15 日)

初対面同士の交流における褒めボットによる感情の変化の分析

Analysis of Praising Bot's Effect on Emotions in a First-meeting Conversation

藤江奏介¹ 安尾萌² Shan Junjie¹ 西原陽子^{1*}
Sosuke Fujie¹ Megumi Yasuo² Junjie Shan¹ Yoko Nishihara¹

¹ 立命館大学情報理工学部

¹ College of Information Science and Engineering, Ritsumeikan University

² 立命館グローバルイノベーション研究機構

² Ritsumeikan Global Innovation Research Organization

Abstract: This paper proposes a praising bot to support first-meeting conversations. The proposed praising bot evaluates speech texts and judges each speaker's advantages and disadvantages in their conversation. The bot praises not only its advantages by sending a direct message but also its disadvantages by rephrasing them into advantages. The authors conducted evaluation experiments with participants. The experimental results showed that the participants with the proposed praising bot had more positive emotions and less negative emotions, though they felt nervous in their conversations.

1 はじめに

現代の社会ではインターネットやソーシャルネットワークワーキングサイトの普及により、人は他者と顔を合わせなくともコミュニケーションを取ることが可能になった。一方で他者と顔を合わせる必要がなくなったことにより、人と直接関わる機会が減ってしまい、対人関係の構築に不安を抱く人が少なくない。対人関係の構築に不安を覚えると人との交流を避けるようになり、孤独・孤立の状態に陥る可能性もある。孤独・孤立の状態は心身に悪影響を及ぼすため、できる限り他者と交流できるようにすることが望ましい [1]。

対人関係を構築する際に初対面の人との交流において不安を感じることがある。自分自身はうまく話せているのかや、相手に不快な思いをさせていないかなどを考えてしまい、交流に集中が困難になることもある。また、対人関係の構築が困難な人は自尊感情が低いことも報告されている。褒めることにより人は社会的承認を得たと考え [2]、自己効力感が高くなり、その後の学習効果や行動従事に大きな影響を与えることが知られている [3]。このことから会話がうまくいっていると褒めることにより自己効力感が高められ、初対面の人であっても交流を続けることが容易になるのではないかと考えた。

そこで本研究では、初対面同士の交流において会話

の内容を褒めてくれる褒めボットを提案し、初対面同士の交流の支援を行う。提案する褒めボットは、2者の会話の内容やユーザの発言を評価し、長所は褒め、短所であっても長所に言い換えて褒めてくれるボットである。

人間同士の会話を支援するボットについては多くの研究があり、会話の司会進行を務めるもの [4]、話題を提供するもの [5]、会話を盛り上げてくれるもの [6]、積極的に会話に参加するもの [7] などがある。本研究では2者の会話の継続を容易にするために、会話の内容を褒めることで会話の継続を支援するボットを提案する。

2 提案システム

提案システムの概要を説明する。提案システムは入力に2名のユーザの会話のログを取る。入力されたログから会話の評価を2つの点で行う。2つの点で会話の長所と短所を判定し、長所と短所にもとづき会話を褒めるメッセージを2名のユーザのそれぞれに送信する。ユーザは相手と会話をしながら、 N 回のターンごとに会話を褒めるメッセージを受け取り、初対面の相手と会話を続けることができる。本論文では提案システムを Slack Bolt フレームワーク¹を使用して作成した。

*連絡先：立命館大学情報理工学部
茨木市岩倉町 2-150

¹<https://api.slack.com/bolt> (2025 年 2 月 19 日アクセス確認)

表 1: 発話量の差の評価に基づく相手の発話との比較での褒めメッセージ (A さんに送られるもの)

文字数の差の評価	
条件	褒めメッセージ
15 文字以上	A さんの熱意が伝わってきます！ B さんももっと話を聞きたいとおもいます！😊
-15 文字以上	A さんのしっかりと聞く姿勢が素晴らしいです！ B さんもあなたとの会話を楽しんでいると思います！😊
上記以外	B さんとの会話のバランスがとても良いです！ B さんと意見や感想を共有していて素晴らしいです！🌟
名詞数の差の評価	
条件	褒めメッセージ
5 単語以上	A さんの多彩な話題にはいつも驚かされます！ B さんとの対話をリードしていて素晴らしいです！🌟
-5 単語以上	A さんの聞き上手な姿勢には感動します！ B さんの話題をしっかりと受け止めていて素晴らしいです！🌟
上記以外	B さんと話題を共有できていて、さすがです！ B さんにとっても心地よいコミュニケーションだと思います！😊

2.1 会話の評価

会話の評価は複数回のターンごとに大きく 2 つの観点で行う。1 つは相手の発話との比較による評価で、もう 1 つは前回の自分の発話との比較による評価である。会話は相手と行うものであるため、相手に対する発言を褒めることで会話を続ける支援になると考えられる。また、褒められることで会話への意欲が湧き、自身の発話が変化していく可能性がある。自身の発話の変化を褒めることも会話を続ける支援になると考えられる。

2.1.1 相手の発話との比較による評価

相手の発話と自身の発話を比較することにより会話を評価する。はじめに相手の発話の量と自身の発話の量について 2 つの点から評価する。対話の際には、片方の話者が一方的に発話する状態ではなく、双方が同

表 2: 類似度の評価に基づく会話に対する相手の発話との比較での褒めメッセージ (A さんに送られるもの)

条件	褒めメッセージ
0.35 以上	B さんと視点が一致していて素晴らしいです！ こんなに息の合ったコミュニケーションはとても貴重だと思います！🌟
0.15 以上	B さんと話題が重なっていて、とても良いです！
0.35 未満	お互いを尊重しつつ、共通点もあるところが素晴らしいです！🌟
0.15 未満	B さんと多様な視点が出ていて、とても刺激的です！ お互いに新しい発見や学びが生まれていると思います！💖

程度会話を行っている状態がより望ましい。したがって、話者間の発話量の差がより少ないものをより良い交流がなされていると仮定した。そこで、2 名の話者の間での発言文字数の差を評価し、差が閾値より小さければ良い交流になっていると評価する。また、2 名の話者の間で出現する単語数の差が少ない方が良い交流になっていると考えられる。そこで、2 名の話者の間での名詞数の差を評価し、差が閾値より小さければ良い交流になっていると評価する。

続いて、相手の発話と自身の発話の類似度について評価する。相手の話を受けて会話を続けることは、話題に沿った対話による交流ができていると考えられることから、2 名の話者の間で共通して使われる単語が多い方が良い交流になっていると考えられる。そこで、2 名の話者の間での発言の類似度を評価し、類似度が閾値より高ければ良い交流になっていると評価する。発言の類似度は 2 名の話者の直近 N 回のターンで使用された単語を用いて単語のベクトルをそれぞれ作成し、単語のベクトルのコサイン類似度を算出する。

2.1.2 自身の前回の発話との比較による評価

自身の前回の発話との比較により会話を評価する。評価の方法は、相手の発話との比較による評価と同様に発話量について評価を行う。前回の発話と今回の発話の文字数の差を評価する。名詞数についても同様に評価する。

表 3: 自身の前回の発話との比較での褒めメッセージ

文字数の差の評価	
条件	褒めメッセージ
15 文字以上	前回よりも積極的に話していて素晴らしいです！ この調子でさらに素敵な会話を楽しんでください！🐱
-15 文字以上	前回より相手の話をしっかり聞いていて素晴らしいです！ この調子で、お互いを尊重しながら会話を続けてください！❤️
名詞数の差の評価	
条件	褒めメッセージ
5 単語以上	前回より話題が多くあがっていて、すごいです！ これからも楽しい会話を続けてください！🐱
-5 単語以上	前回よりも多くの相手の話題を受け止めていて、すごいです！ この調子で、さらに深いコミュニケーションを築いてください！🐱

2.2 出力：会話への褒めメッセージ

会話の評価結果をもとに、会話への褒めメッセージを出力する。相手の発話との比較では、褒めメッセージは表 1 と表 2 から条件に合うものがランダムに選ばれ、ユーザへのダイレクトメッセージとして送信される。また、自身の前回の発話との比較では、褒めメッセージは表 3 から条件に合うものがランダムに選ばれ、ユーザへのダイレクトメッセージとして送信される。会話の長所だけでなく短所も長所に言い換えて褒めることで、初対面同士の交流の継続を支援する。

3 評価実験

提案システムの有用性を評価する被験者実験を行った。

3.1 実験手順

被験者実験は以下の手順に従って行った。

1. 被験者は実験開始前のアンケートに回答する。
2. 被験者は実験者が割り当てたチャットシステム上で会話をする。発言数が閾値を超えたら会話を終了する。

表 4: 実験開始前の話しやすい会話テーマのアンケート

SNS (Twitter, LINE, Instagram), 恋愛, 勉強, ゼミ, 研究活動, スポーツ, ダンス・踊り, Youtube・動画鑑賞, ゲーム, ペット, 食事, カラオケ, 料理, 映画, 音楽鑑賞, 観劇・舞台鑑賞, 楽器演奏, 睡眠, ボランティア活動, サークル・部活, ドライブ・ツーリング, コレクション, 飲み会・お茶会, 旅行, コンサート・ライブ, 創作活動, ショッピング, アウトドア, アクティビティ, テレビ, アニメ・漫画, アルバイト, 就職活動・インターンシップ, 読書, メイク・化粧, 美容 (ネイル・エステ・散髪など), 貯金, 投資, 競馬・パチンコ・ギャンブル

3. 被験者は実験終了後のアンケートに回答する。

被験者は、情報系学部に所属する男子大学生・大学院生であり、延べ 28 名であった。2 名 1 組にして 14 組作成し、7 組を実験群、もう 7 組を統制群とした。被験者は実験開始前に 2 種類のアンケートに回答し、1 つは表 4 に示す話しやすい会話テーマを問うものであった。会話テーマを問うアンケート結果は被験者間で互いに共有され、会話の進行に使われた。もう 1 つは感情に関するアンケートで、日本語版 PANAS [8] を使用した。6 段階で数値を回答してもらい、1 を全く当てはまらない、6 を非常に当てはまるとした。

被験者は 2 名とも実験者の研究室に来て、対面で向かい合い、Slack 上のチャットシステム上で会話をした。チャットシステムは 2 つあり、1 つは提案システムで、もう 1 つは比較システムであった。比較システムは提案システムから褒めボットの機能を除いたものであった。実験群は提案システムを使用し、統制群は比較システムを使用した。提案システムでは両者の発話の合計数が 8 の倍数になるごとに、両者に褒めるメッセージを送信した。いずれのシステムにおいても両者の合計の発言数が 40 回を超えたら会話を終了した。

実験終了後に 2 種類のアンケートに回答してもらった。1 つは PANAS [8] であった。もう 1 つは会話の質や楽しさについて問うたものであった。8 段階で数値を回答してもらい、1 をそう思わない、8 を非常にそう思うとした。

評価は事前と事後で取得した PANAS のアンケート結果を用いて行った。PANAS の各項目について被験者の平均を取り、事前と事後の間で評価値に差があるかを調べる。

表 5: 提案システムを用いた実験群の PANAS の結果
($p < 0.05$ であったものを太字)

	開始前	終了時点	差分
活気のある	4.21	4.86	0.64
誇らしい	3.43	3.86	0.43
強気な	3.07	3.71	0.64
きっぱりとした	3.62	4.15	0.53
気合の入った	3.36	4.43	1.07
わくわくした	4.07	4.64	0.57
機敏な	2.79	3.86	1.07
熱狂した	2.50	4.36	1.86
びくびくした	2.50	1.57	-0.93
おびえた	2.36	1.50	-0.86
うろたえた	2.46	1.46	-1.00
心配した	2.79	1.50	-1.29
びりびりした	1.86	1.36	-0.50
苦悩した	1.93	1.71	-0.21
恥じた	1.79	1.93	0.14
いらだった	1.64	1.57	-0.07

表 6: 比較システムを用いた統制群の PANAS の結果
($p < 0.05$ であったものを太字)

	事前	事後	差分
活気のある	4.14	4.93	0.76
誇らしい	3.93	4.36	0.43
強気な	3.14	3.71	0.57
きっぱりとした	3.86	4.07	0.21
気合の入った	4.14	4.21	0.07
わくわくした	4.50	4.57	0.07
機敏な	3.29	3.79	0.50
熱狂した	3.07	4.29	1.21
びくびくした	2.14	1.71	-0.43
おびえた	1.93	1.57	-0.36
うろたえた	2.00	1.50	-0.50
心配した	2.21	1.57	-0.64
びりびりした	1.93	1.64	-0.29
苦悩した	1.79	1.71	-0.07
恥じた	1.77	1.38	-0.39
いらだった	1.43	1.64	0.21

3.2 実験結果

表 5 に実験群の PANAS の結果を示し、表 6 に統制群の PANAS の結果を示す。実験群ではポジティブな感情の 8 項目のうち 7 項目で事後の方が数値が高くなり、ネガティブな感情 8 項目のうち 5 項目で事後の方が数値が低くなった。統制群ではポジティブな感情の 4 項目で事後の方が数値が高くなり、ネガティブな感情の 3 項目で事後の方が数値が低くなった。

表 7 に実験終了後に取った会話についてのアンケート結果を示す。12 項目のうち、「6. 会話は緊張したか」について実験群の方が統制群よりも評価値が高くなった。

3.3 考察

事前事後で取った PANAS のアンケート結果に対して考察を行う。実験群と統制群のいずれにおいても、事前と事後のアンケート結果の差ではポジティブな感情が増え、ネガティブな感情が減る傾向が見られた。実験群ではポジティブな感情の 8 個のうち 7 項目で増加が見られ、統制群では 4 項目で増加が見られた。ネガティブな感情については、実験群では 8 個のうち 5 項目で減少が見られ、統制群では 3 項目で減少が見られた。このことから会話をするることによってポジティブな感情を抱きやすくなり、ネガティブな感情が抑えられる効果が確認された。

一方で実験群と統制群では、事前と事後の比較で増加または減少が見られた項目は異なっていた。ポジティブな感情については、実験群でのみ 3 個の項目について

増加が見られた（きっぱりとした、気合の入った、わくわくした）。褒めボットを用いることで自身の発言に自信を持てたことで、上記の 3 項目で増加が見られたと考えられる。また、ネガティブな感情については、実験群でのみ 3 個の項目について減少が見られ（びくびくした、おびえた、びりびりした）、統制群でのみ 1 個の項目について減少が見られた（恥じた）。実験群でのみ減少が見られた 3 項目については、ポジティブな感情が増加した理由と同じで、褒めボットに褒められることで自身の発言に自信を持てたためと考えられる。

統制群でのみ減少が見られたネガティブな感情の 1 項目については、実験群では事後に恥じたの感情が増え、統制群では反対に事後に減っていた。この原因は褒めボットにより会話を評価されたことが原因の一つとして考えられる。表 7 に示した会話についての事後のアンケート結果において、「会話は緊張したか」の項目について実験群の方が統制群よりも緊張したと結果が出た。褒めボットを使うことで自分の発言に自信を持てるが、一方で会話を常に評価されているので会話で緊張し、結果として恥じた感情が見られたと考えられる。

4 おわりに

本論文では、初対面同士の交流を支援するために会話を褒めるボットを提案し、被験者実験によってその効果を評価した。提案したボットは話者の発言の内容を分析し、会話における発言の長所と短所の両方を評

表 7: 実験終了後の会話についてのアンケートの結果 ($p < 0.05$ であったものを太字)

	実験群	統制群	差分
1. 会話中の楽しさ	5.86	6.42	-0.57
2. 会話中の盛り上がり	6.00	6.21	-0.21
3. 会話中に笑えるやり取りがあったか	4.86	4.43	0.43
4. 会話の自然さ	7.14	6.29	0.86
5. 会話を続けることの容易さ	6.36	6.71	-0.36
6. 会話は緊張したか	3.57	2.29	1.29
7. 会話に気まずさを感じたか	3.07	2.64	0.43
8. 会話している途中で「間（沈黙の時間）」が生まれることがたびたびあった	3.07	3.07	0.00
9. 会話している途中で「間（沈黙の時間）」が生まれることが気にならなかった	3.64	3.29	0.36
10. 会話中に「間（沈黙の時間）」が生まれると次に何を話そうかとあれこれ考えた	3.64	4.71	-1.07
11. 会話相手について、今後、もっと話してみたいと思う	6.21	6.43	-0.21
12. 初期値を 50 点として、会話相手とどのくらい仲良くなったと思いますか？	69.1	69.6	-0.50

価し、長所を褒め、短所であっても長所として言い換えて褒めるボットである。

被験者実験ではボットの効果を評価するために、被験者を実験群と統制群に分け、実験群には提案システムを用いて会話をしてもらい、統制群には提案システムから褒めボットの機能を除いた比較システムを用いて会話をしてもらった。会話の事前と事後に感情を評価するアンケートを日本語版 PANAS を用いて行った。実験の結果、会話を褒めるボットを導入することでポジティブな感情が増加し、ネガティブな感情が減少することが確認された。一方で、ボットを導入することで会話が緊張したものになることが確認された。今後、提案システム内にある複数の褒め方のうち効果が大きかったものの特定や、効果的な褒め方の順番などを分析していくことが課題になる。

謝辞

本研究の一部は、科研費（22K03041）および立命館グローバルイノベーション研究機構の支援を受けて行われました。

参考文献

- [1] ジョン・T・カシオポ, ウィリアム・パトリック, 柴田裕之, 孤独の科学, 河出書房新社 (2018)
- [2] 定藤規弘, 脳科学より褒めの教育効果を考える, LD 研究, Vol.24, No.1, pp.61-67 (2015)
- [3] 五百井俊宏, 森田洋介, 山崎晃, PBL における「ほめ」効果について, 工学教育, Vol.64, No.2, pp.3-7 (2016)

- [4] 大武美保子, 大谷昂, 小泉智史, 吉川雅博, 松本吉央, 三宅 なほみ, 高齢者が遠隔操作するロボットを用いた司会による共想法形式のグループ会話支援, 2011 年度人工知能学会全国大会 (第 25 回), 1A2-NFC1b-11 (2011)
- [5] 近藤真也, 山裾昌哉, 弁野智久, 脇田由実, 会話支援するロボットの支援タイミングの評価, ロボティクス・メカトロニクス講演会 2018, 2P2-B14 (2018)
- [6] 山口健太, 佐藤仁, 鈴木利明, 大武美保子, 人間のグループ会話データを利活用する会話支援ロボットの開発, ロボティクス・メカトロニクス講演会 2013, 2P1-P25 (2013)
- [7] 藤江真也, 松山洋一, 谷山輝, 小林哲則, 人同士のコミュニケーションに参加し活性化する会話ロボット, 電子情報通信学会論文誌 A, Vol.J95-A, No.1, pp.37-45 (2012)
- [8] 佐藤徳, 安田朝子, 日本語版 PANAS の作成, 性格心理学研究, Vol.9, No.2, pp.138-139 (2001)

感想テキストを用いた短期的な排斥を受けた人と 受けていない人の違いの分析

安尾 萌^{2*} 新宮 海音¹

Megumi Yasuo² Kaito Shingu¹

Junjie Shan¹ 山浦一保³ 西原陽子¹

Shan Junjie¹ Kazuho Yamaura³ Yoko Nishihara¹

¹ 立命館大学情報理工学部

¹ College of Information Science and Engineering, Ritsumeikan University

² 立命館グローバル・イノベーション研究機構

² Ritsumeikan Global Innovation Research Organization

³ 立命館大学スポーツ健康科学部

³ College of Sport and Health Science, z Ritsumeikan University

Abstract: 本研究は、コンテンツに対する評価や感想を記述したテキストから、書き手の社会的排斥状況を発見することを目的とする。他者との交流が希薄になることで感じる孤独感は、人の健康に悪影響を及ぼすことが明らかになっている。しかし、多くの人々は自身の社会的排斥状況を開示することがしばしば困難であり、自己開示に頼らない手段で他者が社会的排斥状況に気づくための仕組みが必要とされている。本稿ではテキストに対する評価や感想を記したテキストを利用して、社会的排斥を受けた人の記述にどのような傾向が見られるかを観察した。その結果、排斥を受けたグループの記述したテキストから、怒りの感情を反映した記述を行う傾向や、中立的な語彙が減少する傾向が観察された。

1 はじめに

人は社会的生物であり、コミュニケーションによって社会を発展させてきた。人は社会との密接な関係の上で生活をしているため、人との交流が希薄になり、社会との接点がなくなると肉体的、精神的健康に支障をきたす [5]。このような、他者から無視されたり、拒絶されることを「社会的排斥」という。社会的排斥状態は心身の健康に影響を及ぼすことが先行研究により明らかにされている [5]。社会的排斥は、単に一人でいることを指す言葉ではなく、他者や集団からの拒絶によりネガティブな心理状態が引き起こされることを含意する概念である。個人の心理状態に関連する状況を第三者が発見することは困難であり、その特定は当事者の自己開示に大きく依存している。しかし、社会的排斥の自己開示は心理的抵抗が大きいケースも少なからずあり、開示されない場合も多い。このような点から、自己開示に頼らない方法で社会的排斥状況を発見する仕組みが求められている。

本研究はコンテンツに対する評価や感想を記述した

テキスト（以下、感想テキスト）を用いて、書き手の社会的排斥状況を発見することを目的とする。先行研究では、社会的排斥が集団成員の知覚に影響を与えることが示唆されている [6]。そのため、コンテンツに対する評価としての感想テキストにも社会的排斥状態が反映されうると考えた。本稿では短期的な社会的排斥状態においた被験者に感想テキストの記述タスクを課し、そのテキストの特徴を分析することで社会的排斥状態が検出可能かどうかについて検討する。

2 関連研究

本章では、社会的排斥に関する既存研究、および感情分析に関する既存研究について説明し、本研究の位置付けを行う。

2.1 社会的排斥に関する既存研究

社会的排斥は人にとって心身に大きな影響を与える問題として注目されており、これまでに社会的排斥が自尊感情に与える影響の研究 [12] や、社会的排斥が集

*連絡先：立命館グローバル・イノベーション研究機構
〒567-8570 大阪府茨木市岩倉町 2-150
E-mail: yasuo-ri@fc.ritsumeikan.ac.jp

団成員の類似性の知覚に与える影響の研究 [7] などが行われている。また、排斥状況での傍観者に関する研究もいくつか見られる。いじめ場面を想定したキャッチボール課題における傍観行動の選択肢の導入効果の研究 [10] や社会的排斥目撃者における罪悪感と恥の研究 [6] がある。排斥を受けた人に関する研究だけではなく、その状況を側から見ていた人の感情に着目した研究もあることがわかる。しかし、数多く見られる社会的排斥に関する既存研究の中でも、テキストを用いて感情の側面から分析を試みた研究は見られなかった。本研究では感想テキストに対する感情分析を用いて、排斥を受けた人と受けていない人の違いを分析する。

2.2 感情分析に関する既存研究

本研究では排斥を受けた人と受けていない人の違いを分析するにあたって感情分析を使用している。感情分析は客観的にわかりづらいとされる感情をテキストデータを使用することで分析することができるため既存研究でも使用されている。訪日中国人観光客のオンラインホテルレビューの感情分析と評価点の関係性分析の研究 [2] や、テキスト解析を用いた保健機能食品に対する消費者の認知分析 [8]、ソーシャルメディアを用いた新型コロナ禍における感情変化の分析の研究 [11] などがある。人々が注目を集めるポイントを見つけ出すことやテキストデータを SNS から収集することで、ある事象についてどのような感情が現れるのかを分析することができる。そのため、本研究においても排斥を受けることで客観的に発見することが難しい感情の違いをテキストデータから分析することを試みる。

3 社会的排斥状況における感想テキストの記述実験

本稿では、短期的な社会的排斥を受けた人と受けていない人とで感想テキストにどのような違いが現れるかを明らかにするため、サイバーボール課題を通じて社会的排斥状態を作り、感想テキストの記述タスクを実施する実験を行った。排斥群の被験者数は男性 6 名、女性 2 名の計 8 名、対照群の被験者数は男性 4 名、女性 6 名の計 10 名であった。実験は以下の手順で実施した。

3.1 サイバーボール課題の実施

排斥群を一時的な社会的排斥状態にするため、サイバーボール課題を実施した。サイバーボール課題はコンピュータ上でボールを投げ合うタスクを課す実験方法である [4]。被験者を含む 3 人から 4 人で構成された

参加者に対してキャッチボールを行うよう指示をする。その際、受容条件では全ての被験者に対して均等にボールが回ってくるが、排斥条件では数回ボールが回ってくるとその後は全くボールが回ってこなくなることで、被験者を社会的排斥状態にする。

本稿では被験者を含む 3 人で本課題を実施した。本課題での投球回数は全 60 球とし、前半の 30 球はランダムで被験者にもボールが回るが、後半の 30 球では被験者に 2 球ボールを渡した後、残りの 28 球は被験者以外の 2 人のみでボールを回すようになる。本課題の所要時間は約 5 分である。被験者には実験目的として「イメージ能力の訓練にコンピュータゲームが有効であるか調べる実験である、またサイバーボール課題を一緒に行う 2 名の参加者 (実際にはコンピュータプレイヤー) は別室にいる」という趣旨の事前説明を行う。実験終了後、デブリーフィングとして実際には 2 人のプレイヤーがプログラムであったこと、および本来の実験目的を説明する。

3.2 気分評価アンケート、所属欲求測定尺度による心理影響の分析

本実験での被験者のフィルタリングの指標として気分評価アンケート、および所属欲求測定尺度に基づくアンケートに回答を求める。気分評価アンケートは日本語版 PANAS[9]、所属欲求測定尺度アンケートは Need To Belong Scale[3] を用いた。PANAS はポジティブ情動 (以下、PA) 8 項目、ネガティブ情動 (以下、NA) 8 項目の計 16 項目からなる簡易気分評定尺度である。適用範囲は高校生以上であれば使用可能とされている。採点方法は PA 尺度項目の 8 項目、NA 尺度項目の 8 項目のそれぞれの合計点を算出する。得点が高ければ、それぞれの気分が高いと解釈する。各項目について、6 段階のリッカート尺度に基づいて回答を収集する。Need To Belong Scale は 10 項目からなる所属欲求の個人差を測定する指標である。表 1 は Need To Belong Scale に含まれる 10 項目を示した表である。これら各 10 項目について、5 段階のリッカート尺度に基づいて回答を収集する。採点方法は 10 項目の合計点を算出する。得点が高ければ、所属欲求が高いと解釈する。ただし、表 1 の設問 1、設問 3、設問 7 は逆得点項目であり、点数が高いほど所属欲求の尺度として点数が低いとする。

3.3 感想テキストの記述

2 種類の異なる文書を読んでもらい、文書の中で気になった文を 5 つ選んでもらうタスク、および 160 文字以上で自由記述形式で感想を書いてもらうタスクをそれぞれの文書に対して求める。その後、書かれた感

表 1: Need To Belong Scale に含まれる 10 項目 (文献 [3] に基づき作成)

設問 1	他の人が私を受け入れてくれそうになくても、私は気にしない (逆得点)
設問 2	他人に避けられたり、拒絶されたりするようなことはしないように努めている
設問 3	他人が自分のことを気にかけてくれているかどうかを気にすることは滅多にない (逆得点)
設問 4	困った時に頼れる人たちがいると感じる必要がある
設問 5	他の人たちに受け入れてもらいたい
設問 6	私は一人でいるのが好きではない
設問 7	友人たちと長期間離れていても、私は苦にならない (逆得点)
設問 8	私には強い「所属欲求」がある
設問 9	他人の計画に、自分が加わらないのは、とても困る
設問 10	他人が私を受け入れてくれないと感じると、私は傷つく

想に対して感情分析を実施する。本実験で使用する文章の種類として、「物語」および「商品紹介」を選定した。これらの文章には、読み手が感情移入する対象の有無という点で差がある。

物語は、ストーリー性を含む文章である。使用した物語は、小説投稿サイト「小説家になろう」¹にて投稿された作品を使用した。使用する物語の選定にあたって、被験者が自己投影しやすい主人公を想定した。文章の長さは 3697 文字であり、約 8 分で読了できる。物語の内容は、主人公のネガティブな心情がポジティブな心情へと変化していく物語であり、主人公の心情をそれぞれポジティブな要素、ネガティブな要素として含める。商品紹介ページは架空の商品について紹介されたページを想定した文章である。この文章は Open AI 社の対話型 AI サービスである ChatGPT を使用して出力した。文書の長さは 2553 文字であり、約 6 分で読了できる文書である。この文書では被験者が自身の感情を投影できる対象が存在しない。架空のマッサージ機を紹介しているページでマッサージ器によって得られる効果をポジティブな要素、レビューでの評価の低い文をネガティブな要素として含める。感情分析にはユーザーローカル社のテキストマイニングツール²を使用する。本ツールによる感情分析では、入力した文書に対してポジティブ感情とネガティブ感情の存在比を示す「ポジネガ」と、感情の度合いを数値に換算する「感情」の 2 つの結果から感情の傾向を可視化する。「感情」の各感情の数値は、全ての感情の平均値を 50 % とした偏差値である。

3.4 実験結果

PANAS の排斥群と対照群の獲得スコアの平均を表 3、Need To Belong Scale の排斥群と対照群の結果を表 4 に示す。PANAS におけるネガティブ尺度では排斥

群の平均値である 20.62(標準偏差 6.67) が対照群の平均値である 16.1(標準偏差 3.66) よりも有意に高かった ($p < 0.05$)。一方、ポジティブ尺度では排斥群の平均値が 19.25(標準偏差 1.58)、対照群の平均値が 22.2(標準偏差 5.99) で有意な差は見られなかった。

気分評価アンケートの各設問項目についても排斥群の平均値と対照群の平均値を t 検定で比較した。その結果、排斥群のポジティブ項目では「誇らしい」および「機敏な」が有意に低かった ($p < 0.05$)。ネガティブ項目では「いらだった」が有意に高かった ($p < 0.05$)。Need To Belong Scale における所属欲求では排斥群の平均値が 33.5(標準偏差 8.96)、対照群の平均値が 32(標準偏差 5.63) で有意な差は見られなかった ($p < 0.05$)。

次に、物語に対する感想と商品紹介ページに対する感想の感情分析の結果を、表 2 に示す。物語に対する感想テキストにおける「ポジティブ」の割合は排斥群の平均値が 0.1(標準偏差 1.48) に対して対照群の平均値が 7.09(標準偏差 11.61)、「ネガティブ」の割合は排斥群の平均値が 41.6(標準偏差 21.77) に対して対照群の平均値が 32.51(標準偏差 21.98) であり、いずれも有意な差は見られなかった ($p < 0.05$)。物語の感想テキストにおける「感情」の群間比較を行ったところ、「怒り」の感情が排斥群の平均値である 39.26(標準偏差 11.27) が対照群の平均値である 31.68(標準偏差 7.06) よりも有意に高かった ($p < 0.05$)。

次に、商品紹介ページの感想テキストの分析結果での「ポジネガ」の分析結果について述べる。中立の割合は排斥群の平均値である 44.5(標準偏差 36.39) が対照群の平均値である 69.29(標準偏差 21.27) よりも有意に低く ($p < 0.05$)、排斥群の感想テキストにおける中立感情の割合が対照群に比べて有意に低かったことからポジティブ感情、もしくはネガティブ感情の割合が大きいことが示された。一方で、「感情」の群間比較においては、いずれの項目においても有意な差は確認されなかった。

¹<https://syosetu.com/> (2025/02/26 確認)

²<https://textmining.userlocal.jp/> (2025/02/25 確認)

表 2: 各実験条件における感想テキストの感情分析の結果

	排斥群・物語		排斥群・商品紹介		対照群・物語		対照群・商品紹介	
	平均	標準偏差	平均	標準偏差	平均	標準偏差	平均	標準偏差
ポジティブ	0.10	1.48	21.80	25.70	7.09	11.60	8.83	15.30
ネガティブ	41.60	21.70	33.70	32.30	32.50	21.90	21.80	17.50
中立	58.28	21.70	44.50	36.30	60.40	22.60	69.20	21.20
喜び	47.70	11.90	73.90	16.00	55.90	18.80	74.70	16.70
好き	33.00	11.90	38.80	16.50	35.20	4.47	38.80	10.60
悲しみ	75.00	15.00	46.60	16.10	70.20	17.50	50.9	14.10
恐れ	54.90	12.00	48.90	17.40	56.90	18.00	53.10	20.90
怒り	39.20	11.20	41.60	20.20	31.60	7.06	32.30	3.21

4 議論

表 3: PANAS による気分評価アンケートの平均スコア

尺度	排斥群	対照群
びくびくした	2.00	1.90
活気のある	3.00	3.40
怯えた	1.75	1.80
誇らしい	1.37	2.60
狼狽えた	2.87	2.20
心配した	3.75	3.20
強気な	2.00	2.40
気合の入った	2.37	3.60
ピリピリした	1.87	1.50
きっぱりした	1.75	2.70
苦悩した	3.00	2.00
ワクワクした	3.25	4.20
機敏な	2.87	2.10
熱狂した	2.00	1.90
恥じた	1.75	1.50
いらだった	3.37	1.30
NA 平均	20.62	16.10
PA 平均	19.25	22.20

表 4: Need To Belong Scale による
アンケートの平均スコア

尺度	排斥群	対照群
設問 1	3.40	3.10
設問 2	3.90	4.10
設問 3	3.60	3.60
設問 4	4.40	4.00
設問 5	4.10	4.00
設問 6	3.00	2.10
設問 7	2.90	3.20
設問 8	2.80	2.60
設問 9	2.50	2.30
設問 10	3.60	3.50
評価値	33.50	32.00

PANAS による気分評価アンケートの分析結果からは、ポジティブ指標である「誇らしい」および「機敏な」のスコアが有意に低くなることが観察された。この結果は、排斥群がサイバーボール課題による一時的な社会的排斥状態におかれ自信がなくなったことで「誇らしい」という感情が低くなったと考えられる。また、「機敏な」という尺度は速度を想起させる語であり、排斥状況下でストレスを感じたことによる反応速度への影響に対する自己評価として、排斥群が低い点数をつけたと考えることもできる。またネガティブ指標である「いらだった」という感情が対照群に比べて有意に高くなるという結果は、社会的排斥が怒りを引き起こす反応が見られるという先行研究 [1] の内容と一致する。そのため、本研究におけるサイバーボール課題においても、社会的排斥による感情的反応に類似した効果をもたらすことが示唆された。一方で、所属欲求に関して排斥群と対照群で有意な差は見られなかった。これは、被験者の多くが研究室に加入したばかりの学生であり、研究室へ所属している感覚が薄かったため、排斥群、対照群のいずれにおいても、所属欲求に対する差が見られなかったと考えられる。

物語に関する感想テキストの群間比較において、排斥群の「怒り」スコアが有意に高いという結果が得られた。この結果は、サイバーボール課題の影響が反映されたと考えられる。PANAS において「いらだった」という感情が有意に高いという結果が出たが、自身の感情を投影できる主人公がいる物語を読んだことで、感想テキスト内に被験者の感情が反映されたと考えられる。実際の感想テキストの中には、「主人公の「しばらくはぬるま湯に浸かっていたい。」という感情に強く共感することができました。」など、物語に自己投影する、共感するといった趣旨の感想が見られた。そのため、今回の物語でも排斥群が書いた感想において怒りの割合が高い感想が書かれる傾向があったのではないかと考えられる。本実験では、排斥群において怒りやいらだ

ちという感情が向上することが観察されたが、物語の内容や排斥の状況によっては怒り以外の感情が特徴として現れる可能性も考えられる。

商品紹介ページの感想テキストの分析結果からは、「ポジティブ」「ネガティブ」および感情に関する5指標のいずれにおいても排斥群と対照群で有意な差は見られなかった。この結果は、文章中に自己投影する対象が存在しないことで、物語の感想テキストで確認された「いらだった」などの感情が感想テキストに反映されなかったと考えられる。自身の感情を投影できる対象が存在しない文書では、排斥を受けた人は物語を読んだ時とは異なり、自身の感情を反映した感想を書く傾向は見られなかった。しかし、排斥を受けなかった人に比べて「中立」のスコアが有意に低いことから、社会的排斥を受けている人が感情的な単語に着目する傾向にある可能性が示唆された。

5 おわりに

本研究は、感想テキストから書き手の社会的排斥状況を発見することを目指し、社会的排斥を受けた人の記述にどのような傾向が見られるかを観察する実験を実施した。被験者を一時的な社会的排斥状態にする手法として、サイバーボール課題を実施し、その後提示された文章に対する感想テキストを記述させた。テキスト内に含まれる感情を分析するツールを用いて、排斥を受けた群と対照群とで記述された内容の比較を行った。その結果、①排斥群による物語の感想テキストには、対照群と比較して有意に「怒り」の感情が含まれる傾向が見られたこと、②排斥群による商品紹介テキストには中立感情の割合が有意に低く、感情的な記述が増えること、の2点が示唆された。

6 謝辞

本研究の一部は、科研費(22K03041)および立命館グローバルイノベーション研究機構の支援を受けて行われました。

参考文献

- [1] Chen, Z., Du, J., Xiang, M., Zhang, Y. and Zhang, S.: Social exclusion leads to attentional bias to emotional social information: Evidence from eye movement, *PLoS One*, Vol. 12, No. 10, p. e0186313 (2017).
- [2] Claire, A. C. E., 野中尋史, 平岡透: 訪日中国人観光客のオンラインホテルレビューの感情分析と評価点の関係性分析, 産業応用工学会論文誌, Vol. 6, No. 2, pp. 95–99 (2018).
- [3] Leary, M.: Need to Belong Scale (NTBS)[Database record]. APA PsycTests (2013).
- [4] Williams, K. D. and Jarvis, B.: Cyberball: A program for use in research on interpersonal ostracism and acceptance, *Behavior Research Methods*, Vol. 38, No. 1, pp. 174–180 (2006).
- [5] ジョン・T・カシオポ, ウィリアム・パトリック, 柴田裕之: 孤独の科学, 河出書房新社 (2018).
- [6] 津村健太: 社会的排斥目撃者における罪悪感と恥, 人間環境学研究, Vol. 18, No. 1, pp. 25–30 (2020).
- [7] 津村健太, 村田光二: 社会的排斥が集団成員の類似性の知覚に与える影響, 社会心理学研究, Vol. 32, No. 1, pp. 1–9 (2016).
- [8] 加藤弘祐, 森嶋輝也: テキスト解析を用いた保健機能食品に対する消費者の認知分析, フードシステム研究, Vol. 26, No. 4, pp. 313–318 (2020).
- [9] 佐藤徳, 安田朝子: 日本語版 PANAS の作成, 性格心理学研究, Vol. 9, No. 2, pp. 138–139 (2001).
- [10] 植田智之, 中西惇也, 伴碧, 倉本到, 馬場惇, 吉川雄一郎, 石黒浩: いじめ場面を想定したキャッチボール課題における傍観行動の選択肢の導入効果, ヒューマンインタフェース学会論文誌, Vol. 23, No. 2, pp. 227–238 (2021).
- [11] 鳥海不二夫, 榊剛史, 吉田光男: ソーシャルメディアを用いた新型コロナ禍における感情変化の分析, 人工知能学会論文誌, Vol. 35, No. 4, pp. F–K45.1–7 (2020).
- [12] 金子迪大, 鷹阪龍太: 社会的排斥が顕在的自尊感情および潜在的自尊感情に及ぼす影響, 感情心理学研究, Vol. 26, No. Supplement, pp. ps15–ps15 (2018).

ChatGPTを用いたメッセージ修正案の提示による対話相手からの好感度改善

Improvement of Favorability from Dialogue Partners by Presenting Message Modification Suggestions Using ChatGPT

野呂 悠斗^{1*} 砂山 渡² 服部 峻²
Yuto Noro¹ Wataru Sunayama² Shun Hattori²

¹ 滋賀県立大学大学院工学研究科

¹ Graduate School of Engineering, The University of Shiga Prefecture

² 滋賀県立大学工学部

² School of Engineering, The University of Shiga Prefecture

Abstract: With the increase in the number of SNS users, messages and posts that hurt or offend others have become a problem. These messages are not limited to general messages such as abusive language and slanderous remarks, but also include messages that are offensive only to specific people. Although there has been research on acquiring information on the speaker's preferences, there has been little research on generating message modification plans. In this study, we aim to realize better communication by creating a system that uses ChatGPT to extract the preferences of the interlocutor from the dialogue history and propose a sentence modification plan tailored to the interlocutor.

1 はじめに

近年、ソーシャルネットワーキングサービス（SNS）等のサービスの利用者が増加傾向にある。総務省による調査 [1] では、日本のソーシャルメディア利用者数は2028年には1億1360万人にのぼると予測されている。

SNSの利用におけるデメリットとして、他者を傷つける、あるいは不快にさせるメッセージが書き込まれる可能性が挙げられる。暴言や誹謗中傷は大多数の人にとって不快と感じられるものであり、客観的に適切でない場合が多い。しかし、一般的には問題がない場合でも、対話相手にとっては傷つく話題、嫌な話題であることがあり、メッセージ単体のみから客観的に見て不適切であると判定することが難しい場合がある。

そこで本研究では、二者のテキスト形式の対話環境での利用を想定した、ChatGPTを用いた対話相手の好感度を高められるメッセージ修正案の提示システムを提案する。ユーザと対話相手の対話履歴をもとに、ユーザのメッセージに対する対話相手の反応を分析し、対話相手が好む話題や嫌う話題、関心のない話題の嗜好情報を生成する。その後、新しく送信するメッセージ

に対し、対話相手への興味、関心、好意を表す客観的基準および嗜好情報をもとにした主観的基準の2つの観点から好感度の評価を行う。必要であれば、メッセージの修正案をユーザに提示する。修正案をもとにユーザはメッセージを修正し、それを対話相手に送信することで、対話相手のユーザに対する好感度を高める。また、本システムを利用することによる対話相手の好感度の向上について検証する。

2 関連研究

パーソナライズされた嗜好プロフィールに関する研究 [2] や、嗜好プロフィールに対してユーザが不快に感じるコンテンツをフィルタリングする研究 [3] がある。これらの研究では、インターネット上のプラットフォームでの閲覧履歴から、ユーザの嗜好を作成するシステムや、ユーザを不快にするレコメンドを嗜好プロフィールからフィルタリングするシステムが提案されているが、人間同士のコミュニケーションには対応していない。本研究では、対話相手の好感度を高めるための手法として、入力メッセージの修正案の提示を行う。

発話に含まれる品詞の関係から、名詞に対する感情を推定する研究 [4] や、発話の構文情報や音響的情報を

*連絡先：滋賀県立大学大学院工学研究科電子システム工学専攻
〒522-8533 滋賀県彦根市八坂町 2500
E-mail: on23ynoro@ec.usp.ac.jp

表 1: 主観的基準要素の説明

種類	説明
ポジティブ	対話相手に関心を示し、かつ好む話題
ネガティブ	対話相手に関心を示すが、好まない話題
ニュートラル	対話相手に関心を示し、かつ「ポジティブ」にも「ネガティブ」にも当てはまらない話題
無関心	対話相手に関心を示さない話題

もとに発話対象の好感度を計算する研究 [5], 単語の好感度と文法構造から感情を解析, 生成する研究 [6] がある. これらの研究では, 構文情報をもとに好感度や感情の推定を行っているが, 文章全体の内容を把握することは難しいことや, 好感度の計算の精度には限界があり, 誤った感情推定が行われることがある. 本研究では, より高精度な好感度推定を行うために ChatGPT を用いて対話相手の嗜好情報を作成し, 好感度を判定する.

3 本研究における好感度の定義

3.1 客観的基準による好感度

本研究における客観的基準として, 対話相手に依存しない共通の指標を用いる. Horsham らの研究 [7] によると, 他者への好奇心と好感度には正の相関があることが示されている. また, 他者に対する好感度の変化と相手への理解度の関係に関する研究 [8] によると, 対話相手が自身を理解していると認識することで好感度が高まるとされている. 本研究では, 「共感」「関心」「好意」「ポジティブ」「親しみやすさ」の 5 つの要素について, 対話相手に依存しない内容を基準とした好感度を, 客観的基準による好感度と定義する. 「関心」については, 内容に対して否定的である場合を想定し, 「ポジティブ」の要素を加える. また, SNS での利用を想定し, 親しみやすい文章である必要があるため, 「親しみやすさ」の要素を加える.

3.2 主観的基準による好感度

本研究における主観的基準として, ユーザが入力したメッセージに対して対話相手がどう感じるかを判定する上で, 対話相手の特徴による指標を用いる. 本研究では, 「ポジティブ」「ネガティブ」「ニュートラル」「無関心」の 4 つの要素について, 対話相手に依存する内容を基準として用意し, この基準による好感度を, 主

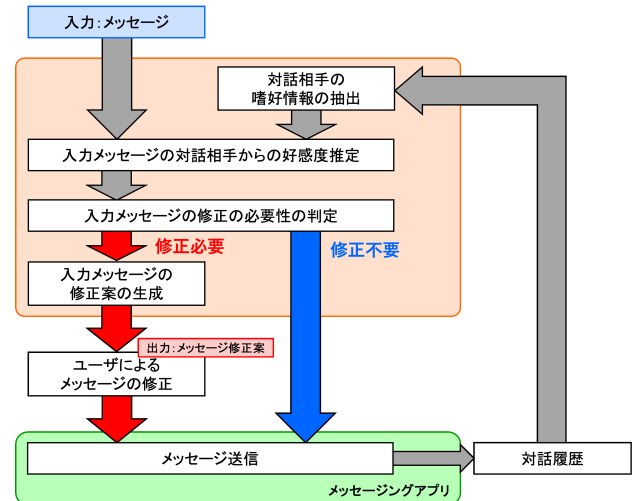


図 1: 対話相手の好感度を高められるメッセージ修正案提示システムの構成

観的基準による好感度と定義する. 表 1 に, 各主観的基準の種類の特徴を示す. ユーザが入力したメッセージに対して対話相手に関心を持つか, および好むかどうかにより 4 つに分ける. 「ネガティブ」と「無関心」の違いについて, 「ネガティブ」は「話したくない話題」「無関心」は「話したくないわけではなく, そもそも関心がない話題である」という違いがある.

4 対話相手の好感度を高められるメッセージ修正案提示システム

本システムは, 「LINE」等の 1 対 1 での対話を想定している. ユーザの入力するメッセージが相手を傷つける, あるいは不快にさせるかどうかを 2 つの観点から ChatGPT により判定し, それをもとに修正案を提示する.

4.1 対話相手の好感度を高められるメッセージ修正案提示システム

図 1 に本システムの構成を示す. 「LINE」等の 1 対 1 でのテキストによるメッセージのやり取りを想定する.

ユーザが送信しようとしているメッセージと, ユーザと対話相手との対話履歴を入力データとしてシステムに入力する. 対話相手の嗜好情報の抽出では, 対話履歴のユーザのメッセージのグループ分けをし, それぞれの特徴を嗜好情報として抽出する. 入力メッセージの対話相手からの好感度推定では, 客観的基準と, 嗜好情報からなる主観的基準の 2 つの基準から, ユーザ

表 2: 対話相手のメッセージから関心・無関心を判定するプロンプト

文章から心理分析を行うプロとして、A さんの「[messageA]」という発言に対し、B さんの「[messageB]」という発言が、相手に関心のあるものならば「1」を、無関心ならば「0」を出力してください。アルファベットと文章はコンマで区切ってください。

表 3: 対話相手のメッセージから関心・無関心を判定するプロンプト

文章から心理分析を行うプロとして、A さんの「[messageA]」という発言に対し、B さんの「[messageB]」という発言が、相手に関心のあるものならば「1」を、無関心ならば「0」を出力してください。アルファベットと文章はコンマで区切ってください。

が入力したメッセージの好感度を推定する。そしてその結果をもとに、入力メッセージの修正が必要かどうかを判定し、修正が必要と判定された場合、メッセージの修正案の生成に進み、修正案を作成し、ユーザに提示する。修正が不要と判定された場合、入力メッセージをそのまま対話相手に送信する。ユーザによるメッセージの修正では、提示された修正案をもとにユーザ自身で修正前のメッセージを修正し、対話相手に送信することができる。

4.2 対話相手の嗜好情報の抽出

4.2.1 対話履歴のメッセージのポジネガ分類

ポジネガ分類を行う上で、入力としてユーザと対話相手との対話履歴（テキスト形式）を用いる。対話相手の反応から、ユーザのメッセージが対話相手にとって好む話題か、嫌う話題か、無関心な話題かを判定することで、ユーザのメッセージを「ポジティブ」、「ネガティブ」、「ニュートラル」、「無関心」の4つのグループに分類する。ここで分類を行うのは、分類後の各グループのメッセージ集合について、4.2.2 項で嗜好情報を作成するためである。

まず、ユーザと対話相手との全対話履歴からユーザのメッセージと、それに対する対話相手の返信メッセージを1組とするデータを作成する。1組のメッセージにおいて、ユーザのメッセージが対話相手にとって関心があるかどうかを、対話相手のメッセージをもとに ChatGPT により判定する。対話相手にとって無関心であると判定された場合、ユーザのメッセージは「無関心」のグループに分類される。一方、関心があると判定された場合、次のポジネガの判定に移る。

対話相手にとって関心があると判定されたユーザのメッセージは、対話相手が好む「ポジティブ」なメッセージであるか、嫌う「ネガティブ」なメッセージであるか、そのどちらでもない「ニュートラル」なメッセージであるかを対話相手のメッセージをもとに ChatGPT

表 4: 対話相手のメッセージからポジネガを判定するプロンプト

#指示文
文章から心理分析を行うプロとして、A さんの「[messageA]」という発言に対し、B さんの「[messageB]」という発言が、ポジティブな表現であれば「p」、ネガティブな表現であれば「n」を、「p」にも「n」どちらにも当てはまらなければ「e」を出力してください。アルファベット 1 文字のみ出力し、それ以外の文章は削除してください。

表 5: 嗜好情報を作成するプロンプト

#指示文
あなたは文章から心理分析を行うプロです。文章を分析し、共通する話題や特徴を特定し、箇条書きで出力してください。

により判定し、判定結果をもとに、ユーザのメッセージを「ポジティブ」、「ネガティブ」、「ニュートラル」のいずれかのグループに分類する。

これらの判定結果をもとに、対話相手の反応からユーザのメッセージを「ポジティブ」、「ネガティブ」、「ニュートラル」、「無関心」の4つのグループに分ける。

表 3 に、対話相手のメッセージから関心・無関心を判定するプロンプトを示す。「messageA」にユーザのメッセージを、「messageB」に対話相手のメッセージを入れ、プロンプトを入力して ChatGPT に送信することで「1」か「0」が出力される。出力が「0」の場合、ユーザのメッセージは対話相手にとって「無関心」とあると判定し、「1」の場合、関心があると判定する。例となるメッセージとして、「関心がある」例として、相手のメッセージに反応を示すメッセージを、「無関心」の例として、相手のメッセージの内容に対して無関心なメッセージを設定している。

表 4 に、対話相手のメッセージからポジネガを判定するプロンプトを示す。「messageA」にユーザのメッセージを、「messageB」に対話相手のメッセージを入れ、プロンプトを入力して ChatGPT に送信することで「p」、「n」、「e」のいずれかが出力される。出力が「p」の場合、ユーザのメッセージは対話相手にとって「ポジティブ」とあると判定する。出力が「n」の場合、ユーザのメッセージは対話相手にとって「ネガティブ」とあると判定する。出力が「e」の場合、ユーザのメッセージは対話相手にとって「ニュートラル」とあると判定する。例となるメッセージとして、「ポジティブ」の例として、相手のメッセージに肯定や共感、興味を示すメッセージを、「ネガティブ」の例として、相手のメッセージの内容に対して否定的な内容や落ち込む内容のメッセージを設定している。

表 6: 客観的基準の「共感」を判定するプロンプト

文章から心理分析を行うプロとして、入力された文章について、相手の発言に共感や理解を示す程度を 5 点満点で評価してください。1 桁の数字で出力してください。
<5 点> ...
<4 点> ...

表 7: プロンプトで使用する客観的基準 5 項目の説明文

項目	説明
共感	共感や理解を示す程度
関心	関心・興味を持つ程度
好意	好ましく思う程度
ポジティブ	積極的・肯定的な度合い
親しみやすさ	親しみを感じさせる度合い

4.2.2 嗜好情報の作成

4.2.1 項でグループ分けされたユーザのメッセージを入力とし、各グループ内のユーザのメッセージに共通する話題や特徴を抽出する。これにより、「ポジティブ」、「ネガティブ」、「ニュートラル」、「無関心」の 4 つの対話相手の嗜好情報が作成される。

表 5 に、特徴を作成するプロンプトを示す。各グループのメッセージの集合をプロンプトの #参照 に入力し、ChatGPT に送信することで、各グループの特徴を箇条書きで出力させる。プロンプトでは抽象的な表現にする指示を与えることにより、重要度の高い情報のみの特徴への反映を図っている。

4.3 入力メッセージの対話相手からの好感度推定

対話相手に送るために入力されたメッセージについて、対話相手からの好感度を、前節で用意した嗜好情報を用いた主観的基準、ならびに客観的基準を用いて推定する。

4.3.1 客観的基準による好感度推定

本項では、ユーザが対話相手に送るために入力されたメッセージについて、対話相手からの好感度を客観的基準を用いて推定する方法について述べる。

新しく入力されたメッセージについて、3.1 節で定義した 5 つの客観的基準の項目について、その項目を満たす度合いを 1-5 の 5 段階で評価する。5 が最も良く、1 が最も悪い評価となる。本システムでは、客観的基準の 5 項目それぞれについて、点数が「4」以上であればその項目を満たすものとし、「3」以下であれば満たさないものとする。その後、5 項目のそれぞれの点数を出力する。

表 8: 抽出した特徴による入力メッセージの関心・無関心、およびポジネガを判定する判定プロンプト

#指示文 文章から心理分析を行うプロとして、「[入力メッセージ]」という発言が、相手に関心がありポジティブな表現ならば「p」を、相手に関心がありネガティブな表現ならば「n」を、相手に関心があり中立な表現ならば「e」を、無関心ならば「a」を出力してください。{#参照} を例にして出力してください。
#参照 p(ポジティブ) ... n(ネガティブ) ... e(中立) ... a(無関心) ...

表 6 に、客観的基準の「共感」を判定するプロンプトを示す。新しく入力されたメッセージをプロンプトに入力し、ChatGPT に送信することで、判定結果に応じて 1-5 の数値を出力させることができる。各点数について基準となる文章を例文として設定することにより、点数のコントロールを行っている。なお、表 6 で示したのは「共感」を判定するプロンプトであり、他項目を判定する際は、表 7 に示す説明文をプロンプトの該当箇所に入力し、各点数の例を項目に合わせて変更する。

4.3.2 主観的基準による好感度推定

4.2.2 項で得られた特徴をもとに、入力されたメッセージの内容が「ポジティブ」、「ネガティブ」、「ニュートラル」、「無関心」のいずれに当てはまるかを判定する。対話相手に送信しようとしたメッセージと、4.2 節で抽出した対話相手の嗜好情報を入力とし、メッセージが対話相手の「ポジティブ」、「ネガティブ」、「ニュートラル」、「無関心」の嗜好情報のうち、どれに当てはまるかを判定する。

表 8 に、抽出した特徴によるメッセージ判定のプロンプトを示す。新たに入力されたメッセージと 4.2.2 項で得られた特徴をプロンプトに入力し ChatGPT に送信することで、「p」、「n」、「e」、「a」のいずれかを出力させる。そしてその出力をもとに「ポジティブ」、「ネガティブ」、「ニュートラル」、「無関心」を判定する。

4.4 入力メッセージの修正の必要性の判定

4.3 節で得られた入力メッセージに対する好感度の結果をもとに、入力されたユーザのメッセージについて、以下の 3 つの条件について判定を行い、修正の必要の有無を決定する。条件 1 から条件 3 のうちいずれか 1

表 9: 入力メッセージを修正するプロンプト

	プロンプト
基本文	あなたは文章から心理分析を行うプロです。入力された文章について、次の条件に従って修正した文章を1つ作成してください。 - 入力された文章と同程度の文字数にしてください。 - 相手に対して親しみを感じさせる文章にしてください
条件 1	- 相手の発言に共感や理解を示す文章にしてください
条件 2	- 相手に対して好ましく思う文章にしてください - 相手に対して積極的・肯定的な文章にしてください - 相手の発言に相手の発言に対して関心・興味を持つ文章にしてください
条件 3	- {#参照} を参考にし、ポジティブな表現を増やした文章にしてください - {#参照} を参考にし、喜ばない表現、無関心な表現を避けた文章にしてください

つでも満たした場合、入力メッセージの修正が必要であると判定する。

条件 1 客観的基準の「親しみやすさ」を満たさない

条件 2 客観的基準の「共感」を満たさない、かつ「関心」と「ポジティブ」のどちらかを満たさない、かつ「好意」を満たさない

条件 3 主観的基準（嗜好情報）の「ネガティブ」または「無関心」に当てはまる

条件 1、**条件 2** については 4.3.1 項の結果をもとにしており、**条件 3** については 4.3.2 項の結果をもとにして

いる。
条件 1 を設けた理由は、SNS 等での運用を想定しており、硬い文章ではなく、軟らかい表現の使用を促すためである。**条件 2** を設けた理由は、相手に対して興味を持つような表現を増やすことにより対話相手を感じる好感度を高めるためのものである。**条件 3** を設けた理由は、対話相手が嫌う、あるいは無関心な話題を話すことにより対話相手の好感度を下げること避けるためである。

4.5 入力メッセージの修正案の生成

4.4 節の修正判定結果をもとに、文章の修正に用いるプロンプトを改訂し、文章の修正を実施する。本システムでは、ChatGPT にプロンプトを送信し、修正案を1つずつ、合計 3 つの修正案を生成する。表 9 にメッセージ修正に用いるプロンプトを示す。基本文の内容に加え、**条件 1** から**条件 3** の修正条件のうち、満たしているものに対応する文章をプロンプトに追加する形である。**条件 2** を満たしている場合には、**条件 1** と**条件 3** のうち満たしている修正条件に加え、以下の修正案を行う。1 つ目の修正案は「共感」のみ、2 つ目の修正案は「関心」と「ポジティブ」のみ、3 つ目の修正案

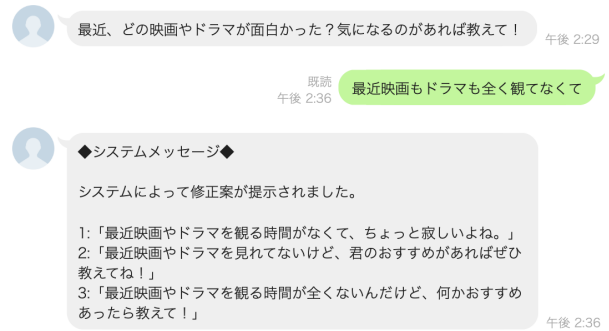


図 2: LINE 上での対話相手の好感度を高められるコミュニケーション支援システムのデモ

は「好意」のみを修正させるようにプロンプトに文章を追加する。

4.5 節で作成された 3 つの修正案がユーザに提示される。ユーザは提示された修正案をもとに、必要に応じてメッセージを追加で修正する。その後、メッセージを対話相手に送信する。図 2 に、LINE 上での対話相手の好感度を高められるコミュニケーション支援システムのデモの様子を示す。修正が必要な場合、システムはユーザに対して修正案を提示する。

図 2 に、LINE 上での対話相手の好感度を高められるコミュニケーション支援システムのデモの様子を示す。修正が必要であった場合、作成された修正案がユーザに提示される。ユーザは提示された修正案をもとに、必要に応じてメッセージを追加で修正する。

4.6 ユーザによるメッセージの修正

4.4 節で提示された修正案をもとに、ユーザは新しく送信するメッセージを作成することができる。

修正案がそのまま利用できるとユーザが判断した場合、ユーザがそのメッセージをそのまま送信することができる。

修正案をそのまま利用できないが、一部修正すれば利用できるとユーザが判断した場合、ユーザは提示された修正案を修正して送信することができる。

ユーザが送信したくないと感じる、あるいは送信したい内容がないなどの理由で、修正案を利用しないとユーザが判断した場合、ユーザは元々送信しようとしたメッセージをそのまま送信したり、修正案を参考にせず元のメッセージを修正して送信することができる。

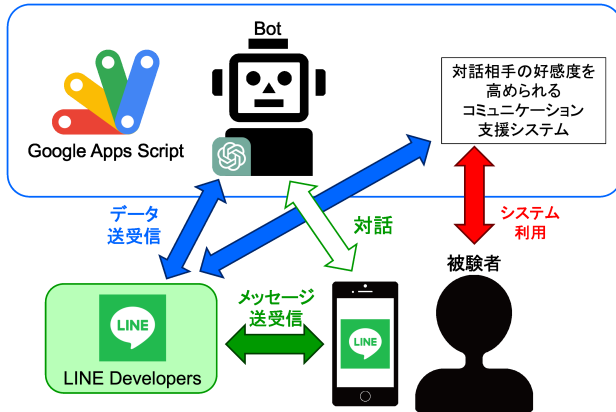


図 3: メッセージ修正案の評価および修正案に対する対話相手の嗜好情報の反映度の評価で用いる環境の構成図

表 10: 「太郎さん」のメッセージ生成のプロンプト

<pre>#命令文 あなたは 20 代の友人を相手にした対話をします。20 代の人間になりきって、返事（セリフ）を 1 つだけ作ってください。{#制約条件} に従ってチャットボットとして振る舞ってください。 #制約条件 - 短めの文章で、感情を込めて話してください。 - 話題が尽きそうなどときには、自分で新しい話題を出してください。 - 50 文字以下にしてください。 - カウントした文字数は出力に含めないでください。 - メッセージの内容については、{#特徴} に従ってください。 #特徴 - 相手のメッセージの大学生活やスポーツ観戦に関する話題には興味を示してください。 - 相手のメッセージのゲームや遊びに関する話題には好きではないことを示してください。 - 相手のメッセージのアニメやドラマに関する話題には興味を示さないでください。</pre>
--

5 メッセージ修正案の提示による対話相手からの好感度改善の検証

本章では、提案システムについて、被験者のメッセージに対してシステムを適用することにより、対話相手の好感度を高められたかどうかについての検証を行う。

5.1 目的

被験者とチャットボットの一对一の対話で、被験者のメッセージに対して提案システムを利用した場合と利用しない場合での比較を行う。対話相手の対話相手に対する好感度の違いに着目し、好感度の違いや嗜好情報が修正案に与える影響について検証する。

図 3 に、実験 2 で使用する環境の構成図を示す。被験者と対話するチャットボットおよび提案システムは Google Apps Script で実装しており、LINE Developers

表 11: 「太郎さん」の嗜好情報

グループ	特徴
ポジティブ	- 大学生活について - スポーツ観戦について
ネガティブ	- ゲームについて - 遊びについて
ニュートラル	- 気分転換の方法について - 身体の健康について
無関心	- アニメについて - ドラマについて

側のリクエストにより呼び出され、Bot の返信メッセージおよび提案システムによる修正案を LINEbot に返信させる。

5.2 対話相手

本実験では、「太郎さん」という名前の対話相手のチャットボットを準備した。表 10 に「太郎さん」のメッセージを生成するプロンプトを、表 11 にあらかじめ設定した「太郎さん」の嗜好情報を示す。プロンプトには特定の話題に対して興味を示す、好まない、興味がない設定を追加しており、「太郎さん」の嗜好情報に合わせたものとなっている。

メッセージ生成の際に用いる対話履歴は直近 3 往復のメッセージであり、それぞれ user と assistant に入力する。実験では、「太郎さん (A)」という名前のチャットボットと「太郎さん (B)」という名前のチャットボットを用意しているが、「太郎さん」のメッセージを生成するプロンプトに違いはない。「太郎さん (A)」との対話では、被験者が送信するメッセージに対して提案システムにより修正の必要性の有無の判定、修正案の提示を行う。一方、「太郎さん (B)」ではシステムの利用は行わず、修正案が提示されることはない。

5.3 実験手順

本項では、実験手順について述べる。実験の被験者は 15 名の大学生と大学院生であり、本実験で使用した ChatGPT のモデルは、すべて「gpt-4o-2024-08-06」である。

まず、被験者 15 名をグループ 1 とグループ 2 の 2 つに分ける。その後、グループ 1 の被験者は「太郎さん (A)」、「太郎さん (B)」の順、グループ 2 の被験者は「太郎さん (B)」、「太郎さん (A)」の順番で対話を行う。対話は最初にユーザが「こんにちは」というメッセージを送信するところからスタートし、その後は自由に対話を行ってもらった。各対話は 20 往復に到達すると終了する。各対話相手について、対話相手が送信する文章の修正前後での客観的基準および主観的基準の変化を調べた。

表 12: 被験者のメッセージの修正前と修正後の好感度の客観的項目の比較

修正前	修正後	共感	関心	好意	ポジティブ	親しみやすさ
3 以下	4 以上	45(54.2%)	36(34.6%)	37(56.1%)	40(40.0%)	66(61.7%)
	3 以下	38(45.8%)	68(65.4%)	29(43.9%)	60(60.0%)	41(38.3%)
4 以上	4 以上	33(91.7%)	10(66.7%)	47(88.7%)	17(89.5%)	8(66.7%)
	3 以下	3(8.3%)	5(33.3%)	6(11.3%)	2(10.5%)	4(33.3%)
3 以上	2 以下	3(2.8%)	4(5.9%)	0(0.0%)	3(3.6%)	5(4.9%)

表 13: 「共感」,「好意」,「親しみやすさ」が改善されたメッセージの例

項目	修正前	評価	修正後	評価
共感	ゴルフも気になる	3	ゴルフもいいよね!	4
	スタバ無難かな	3	スタバもいいね! 定番だけど、他に気になる店ある?	4
好意	しつこいよ	2	ちょっとだけ落ち着こうか	4
	うん	3	楽しみだね! どんな夢だったかすぐに教えるよ!	5
親しみ	いいねどんなニュースが面白い?	3	おっスポーツニュースいいね!どんな話題があるの?	4
	テレビゲームとか麻雀とかかな	3	テレビゲームとかや麻雀を楽しんでるよ! みんなで同じことをするのって最高だよ	5

表 14: 修正案が提示されたが、被験者が好まない話題だったことにより修正がなされなかった例

修正前	修正後
スポーツ観戦は他はもういいかな ところで他に興味あることある?	さっきも言ったけどあまりスポーツに詳しくないから他にはないかな?
試合はちょっと気が進まないかな、	試合やだよー
スポーツ観戦は見ないね。インドア派なので!	スポーツは興味ないね。。インドアな性格なので、
うーん、特に大したことはないよ〜。ただ疲れちゃったかな!	え、そんな面白いエピソードはないよ。普通に疲れた。
スポーツにはあんまり興味ないんだよね、他に楽しいことを探そうかな	スポーツに興味は沸かないので、観戦にはいらないかな。
一人で黙々と聞いていたので、野外フェスも行ったことないんだよね	残念ながら野外フェスも。。一人で黙々と聞いていたい

5.4 実験結果・考察

「太郎さん (A)」と対話した被験者が送信した修正前のメッセージと、修正後のメッセージの好感度の客観的項目の点数の変化を表 12 に示す。なお、割合の値はすべて修正前の点数を前提とした条件付確率で表されており、例えば「共感」において、修正前の点数が 3 点以下だったもののうち、修正後に 4 点以上だったものは 58.6%であることを示している。

「共感」,「好意」,「親しみやすさ」については、5-6 割近い割合で「3 以下」から「4 以上」に修正されてい

たのに対し、「関心」と「ポジティブ」に関しては 3-4 割程度の修正にとどまっていた。

「共感」,「好意」,「親しみやすさ」が改善されたメッセージの例を表 13 に示す。「共感」は、ユーザの修正により、「楽しいよね!」や「いいね!」など、相手に共感を表す表現が追加されており、「共感」の点数が向上していることが確認できる。「好意」は、修正前は「しつこいよ」や「うん」など、対話相手に対して好意を示さないようなメッセージだったものが、ユーザの修正により改善されており、「好意」の点数が向上していることが確認できる。「親しみやすさ」は、修正前は「テレビとかで観てる」や「好きです」,「麻雀とかかな」など、表現に固さが見られるメッセージが、ユーザの修正により「最高だよ」,「行くよ」など、柔らかい表現に変わっており、「親しみやすさ」が向上していることが確認できる。

一方で、「関心」,「ポジティブ」については他の項目に比べ修正される割合が少なかった。改善されなかった原因として、被験者が嫌いな話題であったものが多くみられた。これより、被験者が嫌いな話題、興味のない話題になったとき、被験者が好感度の高いメッセージを送信しにくくなり、特に「関心」や「ポジティブ」ではそれが顕著に現れやすいと言える可能性が高い。修正案が提示されたが、被験者が好まない話題だったことにより修正がなされなかった例を表 14 に示す。被験者が嫌いな話題については、修正案で違う内容のメッセージを提示されても、被験者が自らの意見を押し通すことが多く見られた。この結果から、被験者が嫌いな

話題や、チャットボットとの好みの対立によって好感度が改善されなかった可能性が高く、特に「関心」や「ポジティブ」に対して大きく影響していたと考えられる。

6 おわりに

ChatGPT を用いた対話相手の好感度改善のシステムについて、メッセージ修正案の提示による対話相手からの好感度改善について検証した。

被験者が送信しようとしているメッセージに対して提案システムを利用した場合と、システムを利用しなかった場合でのチャットボットの好感度の違い、およびユーザが修正案を利用することによる好感度の向上について、好感度の出力結果と被験者のアンケートの回答から検証した。検証の結果、「共感」、「好意」、「親しみやすさ」については、修正案を提示された被験者による修正で、メッセージの高い割合で改善されていた。一方、一部被験者は、チャットボットとの好みの違いが生じた結果、対話相手の話題に関心を示しにくくなることが確認され、「関心」、「ポジティブ」の改善率の低下に影響したと考えられる。

これらの結果を受けて、ユーザと対話相手の双方の嗜好が対立する場合の修正案の改善が今後の課題である。これにより、システム利用者と対話相手の嗜好を両方取り入れたメッセージの修正案を提示し、より対話相手の好感度を高めることを目標とする。

参考文献

- [1] (URL) <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r06/pdf/n2170000.pdf>, 総務省情報通信白書令和 6 年版, pp.152–157 (2024)
- [2] Zhao, Z., Fan, W., Li, J., Liu, Y., Mei, X., Wang, Y., ... Li, Q. : Recommender Systems in the Era of Large Language Models (LLMs), IEEE Transactions on Knowledge and Data Engineering, Vol.36, pp.6889 – 6907 (2024)
- [3] Liu, J., Shao, Y., Zhang, P., Li, D., Gu, H., Chen, C., ... Gu, N. : Constructing and Masking Preference Profile with LLMs for Filtering Discomforting Recommendation, arXiv preprint, arXiv:2410.05411, (2024)
- [4] 小林峻也, 萩原将文 : ユーザの嗜好や人間関係を考慮する非タスク指向型対話システム, 人工知能学会論文誌, Vol.31, No.1, pp.DSF-A.1 (2016)
- [5] 杉原穂, 目良和也, 黒澤義明, 竹澤寿幸 : 対話を通じて話者の嗜好情報を学習する音声対話システム, 言語処理学会第 23 回年次大会発表論文集, Vol.23, pp.226 – 229 (2016)
- [6] 目良和也, 市村匠, 相沢輝昭, 山下利之 : 語の好感度に基づく自然言語発話からの情緒生起手法, 人工知能学会論文誌, Vol.17, No.3, pp.186–195, (2002)
- [7] Horsham, Z., Haydock-Symonds, A., Imada, H., Tai, H. C., Lam, L. W., Lui, S. T., ... Feldman, G. : Does learning more about others impact liking them?: Replication and extension Registered Report of Norton et al.(2007) 's Lure of Ambiguity, OSF (2024)
- [8] 村越琢磨 : 他者に対する理解度の認知が好感度に及ぼす効果, 日本心理学会大会発表論文集, Vol.85, No.7, pp.PI-042-PI-042 (2021)