

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

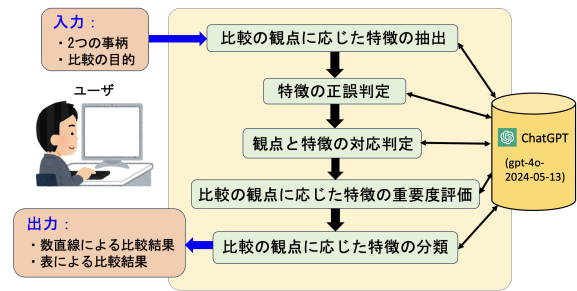


図 1: 2つの事柄の比較支援システムの構成

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、こ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、具体的な内容出力。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 比較の観点到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 比較の観点到応じた特徴の分類

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

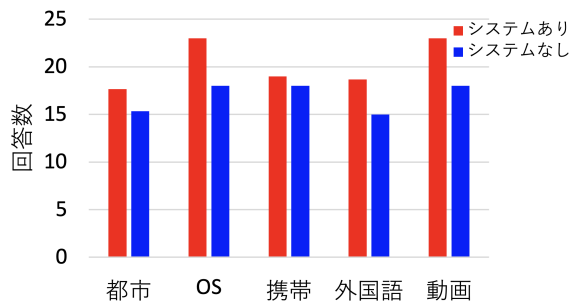


図 3: 比較 1 回あたりの「違い」の回答数 (被験者合計)

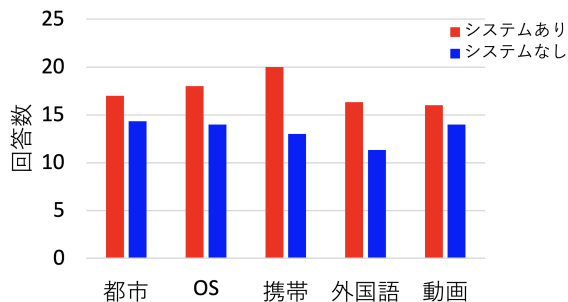


図 4: 比較 1 回あたりの「共通点」の回答数 (被験者合計)

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 becoming 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

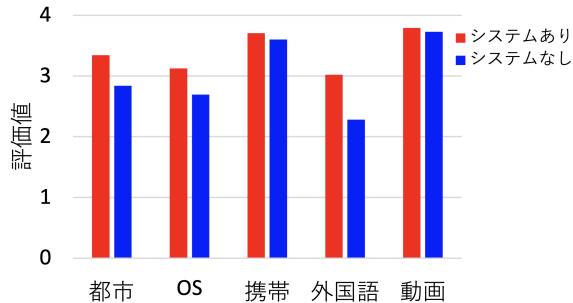


図 5: 「違い」の ChatGPT による評価値 (全回答平均)

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

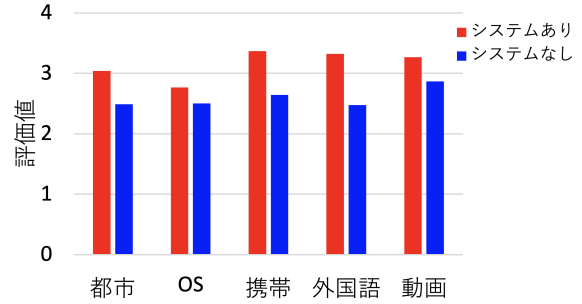


図 6: 「共通点」の ChatGPT による評価値 (全回答平均)

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)