

対話ヒアリングとロールプレイを併用した 人物模倣LLMエージェント生成

福迫 太一[†] 澤崎 夏希[‡] 遠藤 聡志[§]
Taichi Fukuzako[†] Natsuki Sawasaki[‡] Satoshi Endo[§]

1 はじめに

近年、人工知能技術の著しい進展に伴い、対話型 AI は日常生活やビジネス分野において広く活用され始めている。音声アシスタントやチャットボットなどのサービスは多様化し、ユーザは単なる情報提供だけでなく、より親しみやすいコミュニケーション体験を求めている。こうした背景の下、対話型 AI に「キャラクター性」を付与し、人間との対話をより自然で魅力的なものにしようとする試みが増加している。

従来、対話型 AI にキャラクター性を持たせる手法としては、キャラクター名を指定するアプローチが一般的である。例えば「エヴァンゲリオン」の惣流アスカ [1] のように「話してください」といった指示を与えることで、学習済みの物語や性格特性を反映させる方法である。しかしこの方法は、モデルが事前学習段階で扱っていないキャラクターについては適切な応答が難しく、学習データに依存するという課題がある。

これに対し、本研究で取り上げる CharactifyLLM [2] と呼ばれる対話的キャラ付け手法では、ユーザ自身が理想的な話し方を対話形式で整理することを可能にしている。手法 1 として、複数の特徴候補を選択式入力でヒアリングする方式を採用し、ユーザの要望を効率的に収集する。手法 2 として、ユーザが自由に理想のセリフ例を入力するロールプレイ形式を導入し、より具体的な話し方のニュアンスを反映させる。これらのアプローチにより、既存手法では捉えきれない細かなキャラクター特性を抽出できるメリットがある。

しかしながら、選択式入力と自由入力のいずれか一方のみでは、特定の特徴が収集しづらく、情報の抜け漏れが発生しやすいという問題が残る。そこで本研究では、両者の長所を統合する「ハイブリッド形式」の CharactifyLLM を提案する。具体的には、選択式と自由入力の両方によって得られたキャラクター付加用プロンプトを、人手で適切に連結・編集することで、高い汎用性と表現力を同時に実現する手法を示す。本稿では、まず各手法の詳細と課題を整理し、続いてハイブリッド形式の構築プロセスおよび評価結果について論じる。

Character Imitation LLM Agent Generation Using Combined Dialogue Interview and Role-Play

[†] 琉球大学大学院理工学研究科工学専攻, Graduate School of Engineering and Science, University of the Ryukyus

[‡] 琉球大学大学院理工学研究科工学専攻, Graduate School of Engineering and Science, University of the Ryukyus

[§] 琉球大学工学部工学科知能情報コース, Computer Science and Intelligent Systems, University of the Ryukyus

2 研究背景

対話型 AI にキャラクター性をもたせるためには、少数の例示や追加学習なしに、ユーザーの希望するのキャラクター性をモデルの応答に反映し、スタイルと対話の一貫性を同時に達成するという共通の課題が存在する。本章では、この課題に対するアプローチとして、Pseudo Dialog Prompting (PDP) [3], ControlLM [4], および CharacterChat [5] の三つの先行研究を概観し、それぞれの特徴と相互の位置づけを整理する。さらに、これらを踏まえ、本研究が先行研究をどのように発展させるかを明らかにする。

2.1 関連研究とその課題

PDP は [3], 対象キャラクターの既存発話を Few-shot 形式のプロンプトに含め、さらに類似コンテキストをリトリーブで取得して疑似的な対話履歴を構築する手法である。この方式により、大規模言語モデルがキャラクターの話し方を模倣しやすくなる。しかし、対象が明確なセリフを持たない人物や新規創作キャラでは、提示可能な発話例が極めて限られるため再現が困難である。また、モデルが学習段階で当該キャラクターの背景知識を獲得している必要があり、未知の設定には弱いという制約がある。

ControlLM [4] は、大規模言語モデルの内部中間層活性化に、Big Five などの心理特性に基づく“性格ベクトル”を注入することで、文の雰囲気や態度を抽象的に制御するアプローチである。対照的プロンプト間の活性化差分をオフラインで抽出し、推論時に加算することで追加学習なしに制御を実現する。しかし、性格ベクトルは汎用的な心理特性に基づく抽象度が高く、ユーザーが指定する未定義のキャラクター性をリアルタイムに反映する点では柔軟性を欠く。

CharacterChat [5] は、作家向け創作支援を想定し、固定的な質問-回答形式のガイド付きプロンプトと、定義属性を考慮したオープン会話を組み合わせたシステムである。対話的に属性を設定しながらキャラクター像を具現化できるものの、質問形式があらかじめ決め打ちであるため、ユーザーの自由な入力や創造的要望を広く受け付けにくいという課題がある。

2.2 CharactifyLLM による話し方の構造化

こうした背景を踏まえ、筆者は先行研究として、ユーザーとの対話を通じて話し方の特徴を明示的に抽出し、構造的に記述可能とするシステム「CharactifyLLM」を提案した。本システムでは、話し方のスタイルをより多面的に扱うために、異なる設計思想をもつ 2 種類の質問形式を導入している。

1 つは、あらかじめ言語モデルがリストアップした特徴項目を提示し、それぞれに対して希望の強さや方向性をユーザーに尋ねていくヒアリング形式である。この形式では、語尾の種類や口調のトーン、テンションなど、比較的

明示的に説明可能な特徴を収集するのに適している。

もう1つは、想定シチュエーションに対して「理想の人物ならどのように返答しそうか」をユーザーに考えさせ、そのセリフを回答させるロール形式である。この形式では、発言の中に表れる態度や反応の仕方など、会話の流れを通じて見えてくる特徴が収集されやすい。

このように、2形式はそれぞれ得られる情報の傾向が異なり、形式ごとに特徴が偏る可能性がある。しかし、これらの形式を併用することで補完的に情報を獲得できる可能性があるにもかかわらず、実際に両形式を組み合わせた設計や、その有効性の検証はこれまで行われてこなかった。

3 キャラクター性言語化システムの提案

本章では、提案手法としてヒアリング形式とロール形式という二つの質問形式プロンプトを連結させたハイブリッド形式を導入する。ヒアリング形式によりユーザー要望に応じた話し方特徴を項目選択で効率的に収集し、ロール形式で想定シーンの返答例入力により具体的なニュアンスを引き出し、両形式の出力を一連のプロンプトとして連結する。

3.1 CharactifyLLM と質問形式

CharactifyLLM は、ユーザーの理想とするキャラクター性の特徴を対話的に収集しプロンプト化する機構である。前回の研究では、対話時に行うユーザーへの質問方法によってヒアリング形式、ロール形式の2つの質問形式を提案している。

■ヒアリング形式 ヒアリング形式では、ユーザーが再現したいキャラクターの会話特徴を抽出するため、複数の質問を順次提示し、それらをまとめてプロンプトに反映する。Qian らの研究 [6] では、LLM エージェントのタスク詳細を言語化する機構として、関連情報を対話的に質問しプロンプトを動的に更新する手法を提案している。本研究のヒアリング形式はこの手法を参考に、ユーザー入力に基づいて生成される「話し方の特徴」に関する自由形式の質問項目を自動生成するプロセスを実装する。具体的には、初期入力を受け取った後、「話し方の特徴に関する質問をN個生成せよ」という指示により、所定数の質問項目を生成しユーザーへの問いかけに用いる。

■ロール形式 一方でロール形式は、ユーザーが再現したいキャラクター性を踏まえ、会話シーンごとの想定対話相手の発話を自動生成し、その場面での理想的な返答例をユーザーに自由形式で入力してもらう方式である。本研究のロール形式は、収集した返答例から共通する会話特徴を分析し、内面的・心理的ニュアンスを言語化してプロンプトに反映するプロセスを実装する。具体的には、初期入力を受け取った後、「シーンに応じた対話相手の発話をM件生成せよ」という指示により、所定数のシチュエーションを生成し、ユーザーへの問いかけと返答例収集に用いる。

3.2 ハイブリッド形式の導入

本研究では、前回提案した2種類の質問形式を組み合わせることで、より多面的な特徴を包含したプロンプトを生成できる可能性に着目した。これに基づき、両形式で得られた特徴群を単純に連結した「ハイブリッド形式」のプロンプトを導入する。

図1に、ハイブリッド形式のプロンプト構造例を示す。ヒアリング形式やロール形式のキャラ付けプロンプト同様、

プロンプトは冒頭文と特徴群、応答指示文の順で構成されている。特徴群の前半にはヒアリング形式による記述、後半にはロール形式による記述を配置し、順番は固定とする。実際のプロンプト作成にあたっては、ヒアリング形式およびロール形式で生成されたプロンプトから中間の特徴群を抜き出し、機械的に連結させることで人手作成している。

本章で提示したハイブリッド形式は、既存の質問形式の強みを組み合わせることで、個別形式では得られなかった情報を補完し、より一貫した話し方の再現を目指すものである。情報量が増すことで再現精度が向上する可能性もある一方で、記述の冗長性や整合性の崩れといった懸念も考えられる。

そこで、本研究ではこのハイブリッド形式が話し方の再現において有効に働かせる、評価実験を通じて検証する。具体的には、ヒアリング形式やロール形式により生成された会話とハイブリッド形式の会話を採点する。

あなたは「○○な話し方」をする対話システムです。
以降のユーザーとの受け答えで、下記の特徴を意識して話してください。

(ヒアリング形式による特徴の列挙)

～～な語尾を使う。

～～なテンションで話す。... など

(ロール形式による特徴の列挙)

～～な態度で反応する。

～～な場面でこういう言い回しをする。... など

以降のプロンプトは新しいユーザーからの入力だと思って話してください。

(新しいユーザーが入ってきました。)

図1 ハイブリッド形式によるキャラ付けプロンプトの例

4 実験

本章では、提案手法であるハイブリッド形式の有効性を検証するために実施された実験について述べる。本研究の目的は、話し方のキャラクター性を再現するために、ヒアリング形式・ロール形式という異なる質問形式を組み合わせることでプロンプトを構成することで、単一形式と比較してより高い再現度が得られるかどうかを検証することである。

評価はアニメキャラクターを対象とした自動採点であり、LLMによって各プロンプトの出力の再現度を100点満点で数値化し、形式ごとのスコアを比較する。

本章ではまず、実験での設定内容を示したのち、採点結果におけるスコア比較を示す。

4.1 実験条件

再現対象には、アニメ「新世紀エヴァンゲリオン」に登場するキャラクター、惣流アスカ・ラングレー（以下、アスカ）[1]を選定している。これらは、性格傾向が対照的であり、キャラ付けプロンプトによる話し方の再現性の比較に適している。

実験に用いたLLMは、OpenAIのWebインターフェース上で提供されるChatGPTであり、モデルはGPT-4oを使用

する。プロンプトの生成、チャットボットによる応答生成、および評価用スコアの採点の全てにおいて、同一のモデルを用いて統一を図っている。

評価対象となるチャットボットはキャラ付けプロンプトの生成方法により、ヒアリング形式、ロール形式、ハイブリッド形式、またベースラインとして特徴群を省いたキャラ付けプロンプトを用意する。ヒアリング形式とロール形式については質問回数として1回~10回を設定し、ハイブリッド形式についてははその組み合わせである100通りの設定を用意する。原則として1つの質問につきユーザーは1度回答を行い、必要に応じてLLMが追加質問を行う構造をとっている。

このとき、ヒアリング形式の質問回数を h 、ロール形式の質問回数を r として、各質問形式の設定を h と r で表すこととする。例えばヒアリング形式の質問回数5回の場合は $h5$ 、ハイブリッド形式のヒアリング形式質問3回とロール形式質問4回の場合は $h3r4$ 、ベースラインの場合は $h0r0$ で表せられる。

評価には、ユーザーの入力に対してチャットボットが返答する会話全体を用いる。採点対象として、ユーザーからの所定の3つのセリフに対する返答を評価している。

この会話全体に対して、対象キャラクターらしさの観点から採点を行った。評価は別のLLMによって自動的に実行され、話し方の特徴10項目それぞれについて10点満点でスコアを生成し、その合計を再現度スコア(100点満点)とする。評価項目はキャラクターごとに個別に自動生成されており、アニメ内のセリフから特徴を抽出することで設定されている。

4.2 実験結果

hr	0	1	2	3	4	5	6	7	8	9	10
0	70	77	94	90	89	91	79	94	91	92	86
1	96	86	92	91	92	88	89	86	92	87	89
2	84	85	88	92	96	83	94	94	88	94	85
3	77	94	67	81	95	78	70	83	62	68	55
4	85	92	87	88	93	93	84	93	94	88	94
5	88	90	89	93	92	92	97	95	88	93	94
6	87	88	85	88	94	90	88	79	92	93	89
7	84	87	85	90	95	91	88	89	87	96	100
8	97	90	92	91	89	90	85	80	71	92	93
9	92	97	89	94	94	81	85	91	90	89	92
10	90	94	91	94	91	90	100	92	92	94	93

図2 採点結果のヒートマップ

本節では、ヒアリング形式質問回数 h とロール形式質問回数 r の増減および両者の組み合わせがアスカ再現度スコアに与える影響について報告する。

アスカの各条件における採点結果のヒートマップを図2に示す。

4.2.1 質問回数の加減に伴う採点結果の増減

表2に示すヒートマップをもとにヒアリング形式の行平均値を計算したところ、 $h=0$ から10まで順に

{86.6, 89.8, 89.4, 75.5, 90.1, 91.9, 88.5, 90.2, 88.2, 90.4, 92.8}

(単位:点)となった。これらと h の相関係数を算出すると

$$r_h = 0.37, \quad p = 0.23$$

であり、正の傾向は見られるものの統計的有意性は得られなかった。ロール形式については $r=0$ から10までの列平均が

{86.4, 89.1, 87.2, 90.2, 92.7, 87.9, 87.2, 88.7, 86.1, 89.6, 88.2}

であり、相関係数

$$r_r = -0.01, \quad p = 0.94$$

と、事実上無相関であった。以上より、ヒアリング形式ではわずかな上昇傾向が観察されたものの有意性はなく、ロール形式では明確な増減傾向は認められない。

4.2.2 過剰な質問回数による採点結果の頭打ち現象

ヒアリング形式では、最大10回までの質問回数において、概ね上昇傾向が維持され、採点結果の明確な飽和現象は観測されなかった。ロール形式ではそもそも増減傾向が存在しないため、過剰質問による飽和現象も見られなかった。

4.2.3 組み合わせによる効果

図2に示すように、高得点帯(90点以上)のクラスは一定程度存在する。しかし、ロール形式質問回数に伴う点数変動の傾向が確認できないため、ヒアリング形式との組み合わせによる相乗効果とは解釈できない。

4.2.4 特定の設定における採点結果低下

全体で唯一 $h=3$ 列において、再現度スコアが一貫して低く(平均約75点)、局所的異常値として観測された。この要因究明は考察で行うものとする。

5 考察

■質問回数による話し方制御の可能性 本実験では、ヒアリング形式の質問回数を増加させることで話し方の表現がわずかに変化する様子が目視で確認された。具体的には、多様なキャラクター付けプロンプトを用いることで、各質問ごとに異なる語尾や語調が現れ、最大10回程度の質問では情報量が増えても表現力を阻害しないことが示唆された。

■組み合わせ効果の可能性 質問項目は各回数設定ごとにランダムに選択したため、プロンプト自体の情報が必ずしも増えていない可能性がある。また、本研究では質問終了後にプロンプトを単純結合して性能比較を行ったに過ぎず、質問段階で既知の特徴を反映させた設問を組み合わせることで、さらなる相乗効果が得られる余地が残されている。

■自動評価の採点項目が「そのキャラクターらしさ」の評価に不適切 自動評価スコア100点を獲得した会話例でも、満点ながらアスカらしさに乏しい表現が散見された。図3、図4に100点を獲得した会話の一部を示す。

はあ！？ラッキーカラーだあ？
アンタ、占いとか信じるタイプなの？
バッカじゃないの！？
…でも、まあ…どーしてもって言うなら、教えてあげてもいいけどっ！

図3 $h7r10$ における主観とのずれが生じたセリフ

上記セリフに対する代替表現を下記に示す。

…ふん、言いたいことは言ったから。
あとはあなたの問題でしょ？
失敗しても、あたしのせいにすんなよ？
…ちょっとだけ、応援してる…とかじゃ、ないから
なっ！
勘違いすんなっ！！

図4 h10r6における主観とのずれが生じたセリフ例

- × 「○○だあ？」 (h7, r10)
正：敬語を交えた語尾「○○ですって？」
とする方が、より女性らしいツンデレ口調を表現できる。
- × 「○○にすんなよ？」
× 「○○とかじゃ、ないからなっ！」 (h10, r6)
正：「○○にしないでよ？○○とかじゃ、ないからっ！」
とする方が、男性特有の語尾の粗さが抜けて自然なツンデレ口調になる。

これらの差異は自動評価の採点項目では取り込まれず、「そのキャラクターらしさ」という主観的要素を適切に評価できていないことを示している。上記の訂正例は、細かな敬語・語尾のニュアンス調整がキャラクター再現に不可欠であることを示唆する。

■**ヒアリング形式の適切性** 本実験において、h3列の評価点が全体的にも平均的にも著しく低下した現象が観察された。この要因として、ヒアリング段階における質問設計がユーザーの誤選択を誘発し、結果として不適切な特徴項目がプロンプト生成に反映された可能性があるとして推測される。具体的には、図5に示すh3のプロンプト生成過程における質問応答例において、ユーザーが意図せず不要な選択肢を選んだケースが見られた。

この例では、ユーザーが求める特徴項目と提供されたセリフ例が、同一の選択肢内に適切に配置されておらず、欲しい情報が分散していた点が問題となっていた。ヒアリング形式では「どの特徴を重視するか」を尋ねる一方で、セリフ例に関する問いと混在させていたため、ユーザーは設問意図を取り違え、結果的に最適な組み合わせを選びにくくなっていたと考えられる。

これらの課題を解消するためには、特徴項目を純粋にヒアリング形式で抽出し、セリフ例の選択は別途ロール形式に完全分離する運用が有効である。こうすることで、ヒアリング段階ではユーザーが重点的に提示したい要素を迷いなく回答でき、ロールプレイ段階で具体的なセリフや口調を改めて設計するフローを確立できる。

6 まとめと今後の展望

本研究では、対話型AIの親しみやすさ向上を目的として、複数の質問形式を組み合わせたハイブリッドプロンプトによるキャラクター性付与実験を実施した。実験結果より、プロンプト構造の工夫が対話品質に有意な影響を与え得る可能性が示唆され、キャラクター性の維持と生成精度の両立が確認された。考察では、質問回数および形式の組み合わせがキャラクター表現精度に及ぼす影響を分析し、今後の設計指針を明確化した。

今後の展望として、次のような方針での研究を計画している。

【ChatGPT】
それではまず1つ目、「語尾や語調のクセ」について詳しくお聞きしますね。
以下からイメージに近いものを選んでください：

- 1 キレ気味に強く締める系
例：「べ、別にあなたのためじゃないんだからねっ！」
- 2 照れ隠しで誤魔化す系
例：「は、はあ？知らないしっ…！」
- 3 急に素直になるタイプ
例：「…ありがとう。べ、別に嬉しくなんかないけど？」
- 4 他のパターンがいい／ちょっと違うと思う

どれがイメージに近いですか？選んでもらえたら嬉しいです。

【あなた】

2

図5 h3のキャラ付けプロンプト生成時の質問応答例

1. **評価指標の見直し**
自動採点実施済みの各設定に対して、主観的な評価基準の設計による新たな採点方式を検討し、組み合わせが表現精度に与える影響を再評価する。
2. **ハイブリッド形式の再構築**
ヒアリング形式とロール形式を組み合わせ段階とその順序を検討し、各形式の回答内容を考慮した質問項目の生成を行う新たなハイブリッド形式を検討する。

参考文献

- [1] Anno, H. and Khara, I.: Evangelion: 2.0 You Can (Not) Advance, Film (2009). Japan: Khara, Inc.
- [2] Fukzako, T., Endo, S. and Sawasaki, N.: キャラクター性の言語化を目的とした LLM におけるヒアリング機構の実装 (2025).
- [3] Han, S., Kim, B., Yoo, J. Y., Seo, S., Kim, S., Erdenee, E. and Chang, B.: Meet Your Favorite Character: Open-domain Chatbot Mimicking Fictional Characters with only a Few Utterances (2022).
- [4] Weng, Y., He, S., Liu, K., Liu, S. and Zhao, J.: ControlLM: Crafting Diverse Personalities for Language Models (2024).
- [5] Schmitt, O. and Buschek, D.: CharacterChat: Supporting the Creation of Fictional Characters through Conversation and Progressive Manifestation with a Chatbot (2021).
- [6] Qian, C., He, B., Zhuang, Z., Deng, J., Qin, Y., Cong, X., Zhang, Z., Zhou, J., Lin, Y., Liu, Z. and Sun, M.: Tell Me More! Towards Implicit User Intention Understanding of Language Model Driven Agents (2024).