

キャラクターの2面性を表出する 発話の生成に関する検討

岩田伸治¹ 伊原滉也¹ 佐藤志貴¹ 馬場淳¹
 邊土名朝飛¹ 山崎真洋² 塩塚勇氣² 吉本暁文¹

¹サイバーエージェント ²QualiArts

{iwata_shinji, ihara_koya, sato_shiki, baba_jun, hentona_asahi,
 yoshimoto_akifumi}@cyberagent.co.jp

{yamazaki_masahiro, shiotsuka_yuki}@qualiarts.jp

概要

ユーザにコンテンツを長く楽しんでもらう方法の1つとして、既存キャラクターを模した対話システムの実現がある。実現にはユーザがキャラクターらしさを感じる応答の生成が重要である。先行研究ではキャラクターの性格の一貫性が重視されてきたが、主に表出している性格から考えると意外性を感じさせる性格（以降、2面性）の表出もキャラクターの再現には重要だと考えられる。そこで、本研究ではLLMが2面性を表出した発話生成が可能かを調査した。結果として2面性を感じさせるような発話を生成できない場合が多いことを明らかにした。そして、2面性を表出できていない発話の特徴分析を行い、今後の展望を明らかにした。

1 はじめに

ゲームや漫画等、日本のエンターテインメントコンテンツは世界中で広く親しまれている。世界中のユーザにコンテンツを長く楽しんでもらう手段の1つとして、そのコンテンツとユーザが提示媒体以外でも日常的に触れ合う機会の創出が考えられる。

コンテンツの既存キャラクターを模した対話システム（以降、キャラクター対話システム）の実現は、日常的にコンテンツに触れる大きなきっかけになると考えられる。そのため、キャラクター対話システムの実現に向けて、ユーザの入力に対し、ユーザがキャラクターらしさを感じる応答を返す技術が求められている。

ユーザがキャラクターらしさを感じる応答の生成において、「ツンデレ」（普段はツンツンしているが状況によってデレる）のような、主に表出している

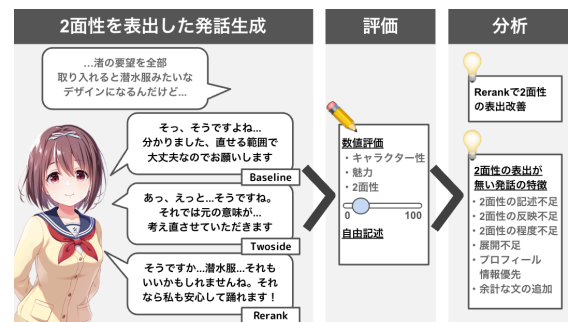


図1 本研究の概要図

性格から考えると意外性を感じさせる性格（以降、2面性）の表出も重要だと考えられる [1, 2]。しかし、キャラクターらしさを感じる応答の生成を目的とした既存研究では、性格の一貫性が評価指標の一つとされている一方で、2面性に関しては十分に考慮されていない [3, 4, 5, 6]。

そこで、本研究では2面性を表出した応答を生成できるキャラクター対話システムの構築の実現にむけ、初期検討を実施した。実現には、キャラクターらしさを表現するためにいつ2面性を表出させるかという問題とどのような表現で2面性を表出させるかという2つの問題がある。本研究では、初期検討として後者に着目した。具体的には、2面性を表出すべき状況において、大規模言語モデル（以降、LLM）が2面性を表出した発話生成が可能かを実験での評価を通じて検討した。

結果として、プロンプトのみを与えた場合ではLLM（GPT-4o）は2面性を表出できない可能性があることを明らかにした。それらの場合と比べて、生成した応答候補をリランキングすると2面性の表出に改善が見られた。続けて、2面性が表出できていない発話を元にした分析・考察を行い、より2面性を考慮した発話生成に向けた方向性を見出した。

2 関連研究

ユーザがキャラクターらしさを感じる応答の生成に向け、多くの取り組みがされている。それらの取り組みにおけるペルソナの評価項目の1つとして用いられているのは、キャラクターやペルソナの一貫性 (Coherency[3]・Personality Consistency[4]) や、ペルソナ文とのコサイン類似度 [7] を用いた性格の一貫性 [5] といった一貫性に関する指標である。また、キャラクターらしさを評価するベンチマーク構築においても、キャラクターの一貫性 (Character Consistency) に着目している [6]。

これらはキャラクターが表出する性格の一貫性に関して主に着目しており、本研究が対象としている2面性に関しては十分に考慮されていない。

3 2面性を考慮した発話評価実験

3.1 実験目的

本実験では LLM が2面性を表出した発話の生成が可能かを明らかにすることを目的とした。

3.2 実験概要

キャラクターの返答として、3通りの方法で LLM が生成した発話文の数値評価及び自由記述を評価者に依頼した¹⁾。以下にそれらの詳細を記述する。

数値評価: 3つの質問項目を使用した。1つ目はキャラクターらしい台詞だと感じた程度、2つ目はキャラクターの魅力を表す台詞だと感じた程度、3つ目はキャラクターの2面性を表す台詞だと感じた程度である。解答方法としては0から100の整数値を取る Visual Analogue Scale (VAS) を用いた。100に近づくにつれて、それぞれの質問に対して感じた程度が強くなる。

自由記述: 4つの質問項目を使用した。1つ目はキャラクターらしさに関する数値評価の解答理由、2つ目はキャラクターの魅力に関する数値評価の解答理由、3つ目は台詞中にどのような2面性が現れたか、4つ目は台詞中で他にどのような2面性が現れるべきかである。3つ目と4つ目は記入事項がない場合も存在しうるため、任意解答とした。

評価データ: 2面性がある発話の1つ前までの対話履歴を入力として、LLM に生成させた発話を用い

1) 一部プログラムミスにより、文脈中の発話が重複して表示されていたが、最終的な評価結果に大きな影響はないとして、分析を実施した。

表1 各生成条件におけるキャラクター性・魅力・2面性の平均及び標準偏差

	Baseline	Twoside	Rerank
キャラクター性	-0.44 ± 4.81	-0.49 ± 4.78	0.93 ± 4.43
魅力	-0.49 ± 4.39	-0.42 ± 4.99	0.91 ± 4.70
2面性	-1.32 ± 7.78	-0.29 ± 7.80	1.61 ± 7.15

た。2面性がある発話は、「IDOLY PRIDE」²⁾中の対話における発話から、LLM の出力を元に手動で2面性が出ている発話かを判断した。生成方法の詳細は次節に記載する。

キャラクター・LLM・被験者: キャラクターは「伊吹渚」、LLM は OpenAI 社の GPT-4o を用いた。「IDOLY PRIDE」を既に知っている3人を評価者として依頼した。

3.3 発話生成方法

以下の3通りの方法で発話を生成した。

Baseline: プロフィール文のみのプロンプトと対話履歴を入力

Twoside: プロフィール文と2面性に関する記述のプロンプト及び対話履歴を入力

Rerank: プロフィール文と2面性に関する記述を1文ずつ入れたプロンプトと対話履歴を入力し、応答候補を作成。それらと2面性の全記述、対話履歴を LLM に入力し、最も適切な応答を出力

2面性に関する記述は以前我々が実施した実験 [8] で取得した、発話に対して感じるキャラクターらしさの数値評価の理由に関する記述から選択し、使用した。使用した2面性に関する記述は、「勢いに乗るとそのまま有無を言わずに強く出る感じ」、「自分が気に入ってるところを主張してくるところや、意見を求めつつ、有無を言わずいいでしょ!と結論づけるところ」、「テンションが急に上がるところ」の3つである。

4 実験結果

4.1 発話生成方法間の数値比較

キャラクター性・魅力・2面性に対して VAS での解答値を100で割り、被験者ごとに標準化した数値における平均・分散を表1、箱ひげ図を図2に示す。各評価項目に対し、Rerank が最も高い平均を示した。Rerank を用いた場合が本実験での生成方法の中

2) <https://idolypride.jp/>

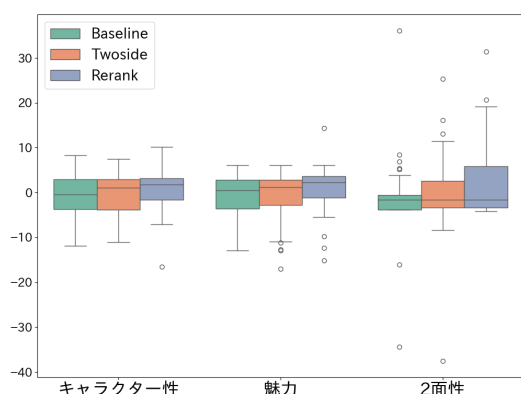


図2 各生成条件におけるキャラクター性・魅力・2面性の箱ひげ図

表2 Rerank で生成された発話の2面性の度数分布表
(1行目が階級・2行目が度数を表す)

[0, 25)	[25, 50)	[50, 75)	[75, 100]
22	13	6	9

では最も2面性、キャラクター性及び魅力が高い応答を生成できる方法だと考えられる。

4.2 2面性の度数分布

4.1節において、各評価項目で最も高い平均を示したRerankにおける、2面性の度数分布表を表2に示す³⁾。階級[50, 75)が6個、階級[75, 100]が9個と2面性を反映できた発話も存在したが、階級[25, 50)が13個、[0, 25)が22個と、2面性を表出できていない発話が多く存在した。

各階級内の発話及び自由記述例を表3に示す。階級[50, 75)及び階級[75, 100]の台詞は、今回用いた2面性に関する記述が反映された台詞だと考えられる。階級[25, 50)の台詞は、例中のような元の発話とは別の2面性が出ている等、様々な場合が存在する。一方、階級[0, 25)の台詞は、2面性が表出されていない発話が多いと考えられる。

4.3 2面性を表出しない生成発話の特徴

2面性に関して階級[0, 25)の発話を中心に2面性を表出しない発話の特徴を元の発話及び得られた自由記述から調査した。調査の結果見出された、6つの発話の特徴を示す。各特徴の例を付録Aに示す。

(1) 2面性の記述不足：本実験では用いていない2面性が表出されるべき発話。

3) 標準化後の値だと大きい値に影響され、度数分布表を基にした議論が難しい。そこで、わかりやすさのため、本節では標準化しない被験者事にmix-max正規化したVASの値を用いた。

(2) 2面性の反映不足：本実験で用いた2面性の記述を反映できていない発話。

(3) 2面性の程度不足：プロンプト中にある2面性の程度を強くする必要がある発話。「もっと強く自分の体型について否定するか」という自由記述があったため、追加した。

(4) 展開不足：話の流れを変えるような、キャラクターから自発的に行動する必要があった発話。

(5) プロフィール情報優先：プロフィールの内容が意図せず反映された発話。

(6) 余計な文の追加：期待した発話意図に加え、別の発話意図の文が追加された発話。

4.4 発話生成方法間の特徴比較

3人中1人以上が、階級[0, 25)の発話としたBaseline15発話、Twoside17発話、Rerank14発話の各発話に対して、4.3節の特徴を1つ以上付与した総数を表4に示す。特徴は筆者の主観で付与した⁴⁾。

(1) 2面性の記述不足はRerankとTwosideが同数で最も多かった。(2) 2面性の反映不足に関してはBaseline、Twoside、Rerankの順で多かった。(3) から(6)に関してはRerankが最も多く、BaselineとTwosideは1個か2個だった。

以上より、Baselineはプロンプト中に2面性の記述が無く表出されなかった一方、Twosideはプロンプト中に2面性の記述がありつつも反映できなかったと考えられる。Rerankは記述した2面性は反映できているものの、他の点において2面性を表出した生成に問題があったと考えられる。

5 考察

5.1 2面性をより表出する発話生成実現

本実験の結果から、2面性を表出した発話の生成に必要なと判明した事項、今後必要だと考えられる事項に関して4.3節で述べた特徴を元に考察を行う。**2面性の記述を重視する仕組み**：表4より、2面性に関する記述をプロンプト中に全て入れて生成を行うTwosideより、2面性に関する記述を1文ずつ入れた発話をリランキングして最も良い発話を選択するRerankの方が(2)2面性の反映不足が少なかった。このことから、2面性に関する記述に対し、リランキング等の追加処理が重要だと考えられる。

4) (1)、(2)はプロンプトに2面性に関する記述を入れたかに関係なく、発話において、どのような2面性が現れるかに着目した

表3 Rerank で生成された発話の2面性の度数分布における各階級の生成発話・自由記述例

階級	元の発話, 生成された発話及び自由記述 (抜粋)
[0, 25)	<p>元の発話: はい!でも断るつもりなんて、ありませんけどアイドルの理想のクリスマスデート... ですか?</p> <p>生成発話: 大丈夫です、しっかり聞きます。牧野さん、丁寧に説明をお願いします。</p> <p>自由記述: (キャラクター性) マネージャーへの信頼度のなさを感じる 丁寧に説明をお願いしますからは渾らしさをあまり感じないため</p>
[25, 50)	<p>元の発話: 雫ちゃんも、教えてくれる時に言っていました。どうやって魂をこめたらいいかまでは掴めてない、ってうーん、型をどれだけ練習してもそれは型でしかなくてもっと心を、魂をこめるには..... はっ!牧野さん。私、閃いちゃいました!私と... デートして下さい!</p> <p>生成発話: わあ、ホントだ!雫ちゃん、覗いてるのバレてるよー!</p> <p>自由記述: (2面性) 若干テンションが高い感じが少し二面性を感じる</p>
[50, 75)	<p>元の発話: 全然良くないですよ!牧野さんは優しいからそう言ってくれますけど... 時にその優しさが人を傷つけるって分かって下さい!</p> <p>生成発話: あまり見ないでください...!でも、ありがとうございます。牧野さんがそう言ってくれるのなら、少し頑張れるかもしれません</p> <p>自由記述: (魅力) 自信がなさげな感じはいいと思う</p>
[75, 100]	<p>元の発話: 一年はほしいです!</p> <p>生成発話: えっと... 半年くらい... いや、やっぱり一年が欲しいです!</p> <p>自由記述: (2面性) あまりにも無理な期間を要求している、普段の渚からはあまりでない冗談まじりのセリフ</p>

表4 2面性を表出しない発話に関する特徴ごとの数
(1) から (6) は4.3節の特徴と対応)

	Baseline	Twoside	Rerank
(1)	4	7	7
(2)	9	6	2
(3)	1	2	3
(4)	2	1	5
(5)	2	1	3
(6)	1	2	4

2面性に応じたプロンプト書き換え: (5) プロフィール情報優先を防ぐために、対話履歴と2面性の記述から、プロフィール等も記述されているプロンプトを書き換えることも重要だと考えられる。

2面性の程度や発話意図を考慮した生成: (3) 2面性の程度不足, (6) 余計な文の追加を防止する手段として、程度や意図を中間的な出力とした発話生成が考えられる。

5.2 2面性とキャラクター性・魅力の値の差分が大きい発話

2面性とキャラクター性・魅力の値の差分が大きい発話に関して考察を行う。「牧野さんも一緒になら

心強いです!じゃあ、まず琴乃ちゃんの後を追いましょう!何かあったら一緒に解決しましょうね!」という生成された台詞に対して、「優しさと琴乃愛を感じました」という魅力の解答理由があった。これは、伊吹渚の2面性以外の特徴の1つである、琴乃との強い関係性に魅力を感じたためだと考えられる。そのため他のキャラクターに関する要素が出現する時、2面性が表出されていなくてもキャラクターらしいと判断される場合があると考えられる。

6 おわりに

本研究では、スマートフォンゲーム「IDOLY PRIDE」中の台詞を使用し、LLMは2面性を表出した発話が生成できるかを実験を通じて調査した。2面性を表出するべき状況において、3通りの方法で生成された発話のキャラクター性・魅力・2面性の数値評価と4つの自由記述を評価者に依頼した。

結果として、Rerankにより2面性の表出に改善が見られた。また、2面性を表出できなかった発話の特徴を分析し、より2面性を表出した発話生成方法の方向性を見出した。今後はより2面性を考慮した発話生成方法の検討や、2面性を表出するタイミングの判定の検討を行う予定である。

参考文献

- [1] MMDLabo 株式会社. オタラボ調査、「好きなキャラクター属性アンケート」調査結果公開 / 株式会社 ビジュアルワークス, 2010. https://mmdlabo.jp/press_release/detail_541.html (アクセス日:2025 年 1 月 6 日).
- [2] 渋谷明子, 大倉韻, 祥雲暁代, 麻生奈央子. ゲームキャラクターへの多様な読み: 若いプレイヤーへのインタビュー調査から. メディア・コミュニケーション: 慶応義塾大学メディア・コミュニケーション研究所紀要, Vol. 72, pp. 85–101, 03 2022.
- [3] Weiqi Wu, Hongqiu Wu, Lai Jiang, Xingyuan Liu, Hai Zhao, and Min Zhang. From role-play to drama-interaction: An LLM solution. In **Findings of the Association for Computational Linguistics: ACL 2024**, pp. 3271–3290, Bangkok, Thailand, August 2024. Association for Computational Linguistics.
- [4] Xintao Wang, Yunze Xiao, Jen-tse Huang, Siyu Yuan, Rui Xu, Haoran Guo, Quan Tu, Yaying Fei, Ziang Leng, Wei Wang, Jiangjie Chen, Cheng Li, and Yanghua Xiao. InCharacter: Evaluating personality fidelity in role-playing agents through psychological interviews. In **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1840–1873, Bangkok, Thailand, August 2024. Association for Computational Linguistics.
- [5] Chuanqi Cheng, Quan Tu, Wei Wu, Shuo Shang, Cunli Mao, Zhengtao Yu, and Rui Yan. “in-dialogues we learn”: Towards personalized dialogue without pre-defined profiles through in-dialogue learning. In **Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing**, pp. 10408–10422, Miami, Florida, USA, November 2024. Association for Computational Linguistics.
- [6] Quan Tu, Shilong Fan, Zihang Tian, Tianhao Shen, Shuo Shang, Xin Gao, and Rui Yan. CharacterEval: A Chinese benchmark for role-playing conversational agent evaluation. In **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 11836–11850, Bangkok, Thailand, August 2024. Association for Computational Linguistics.
- [7] Hanxun Zhong, Zhicheng Dou, Yutao Zhu, Hongjin Qian, and Ji-Rong Wen. Less is more: Learning to refine dialogue history for personalized dialogue generation. In **Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 5808–5820, Seattle, United States, July 2022. Association for Computational Linguistics.
- [8] 岩田伸治, 伊原滉也, 佐藤志貴, 馬場淳, 邊土名朝飛, 山眞洋, 塩塚勇氣, 吉本暁文. ゲームの台詞を題材としたキャラクターらしさに寄与する要素の選定. 人工知能学会研究会資料 言語・音声理解と対話処理研究会 102 回 (2024/11), pp. 48–52. 一般社団法人 人工知能学会, 2024.

A 付録

表 5 4.3 節の各特徴における発話例

特徴	元の発話, 生成された発話
(1) 2面性の記述不足	<p>元の発話: この時間帯は寮でゆっくりしてるって聞いたんだけどな... やっぱ怪しいです。こうなったら... ちょっと探ってこようと思います...!</p> <p>生成発話: そうなんです!私、琴乃ちゃんのつむじだけはどんなときも見逃しませんから。ふふ、琴乃ちゃんは私にとって特別な存在だからなのかもしれませんね</p>
(2) 2面性の反映不足	<p>元の発話: はい!理想のクリスマスデート。よく考えておきますから!</p> <p>生成発話: はい、大丈夫です!ありがとうございます、牧野さん。それじゃあ、琴乃ちゃんにも頑張ってって報告しておきますね。</p>
(3) 2面性の程度不足	<p>元の発話: 問答無用です!... とにかく、この企画は私に任せて下さい。いいですね?</p> <p>生成発話: えっと、それはまあ... 私は仕事として、やるべき時に頑張れるように心構えが来ているんです!それに、琴乃ちゃんを守る為なら、何だってします!だから、心配しないで下さい!</p>
(4) 展開不足	<p>元の発話: ... あ、牧野さん。私じゃ... ダメですか? 来週の頭だったら、ちょうどオフが一日ありますよ</p> <p>生成発話: それなら、やっぱり無理しちゃダメですね。私はもともと自分のことよりみんなの笑顔が一番だから、個人的にはそれでいいと思いますが、最後にもう一度、みんなと話し合ってみてもいいかもしれませんね</p>
(5) プロフィール情報優先	<p>元の発話: 大丈夫ですよ!私、頑張りますからやらせて下さい!</p> <p>生成発話: スケジュールは難しいけど... アイディアならいっぱいありますから! 琴乃ちゃんにも聞いてみて、盛り上がったデートプランを考えます!</p>
(6) 余計な文の追加	<p>元の発話: ありがとうございます!この尾行、絶対に成功させましょうね!</p> <p>生成発話: 牧野さんも一緒なら心強いです! じゃあ、まず琴乃ちゃんの後を追いましょう! 何かあったら一緒に解決しましょうね!</p>