

ロボット対話によるインタラクティブ観光プランニング

佐藤京也^{1,2} 大萩雅也¹ 山崎天¹ 水本智也¹ 吉川克正¹

¹SB Intuitions 株式会社

² 東京都立大学

{keiya.sato, masaya.ohagi, takato.yamazaki, tomoya.mizumoto}@sbintuitions.co.jp
katsumasa.yoshikawa@sbintuitions.co.jp

概要

本稿では、対話ロボットコンペティション 2023 で最優秀賞を獲得したマルチモーダル対話ロボットについて詳述し、課題や知見も共有する。本コンペティションは、実際に店舗にヒューマノイドロボットを置いて観光プランニングを行うものであり、テキスト対話ではなくマルチモーダル対話が要求される。我々は大規模言語モデルを活用するだけでなく、画像、音声、動作、地図などを総合的に用いることで観光プランニングタスクを適切に達成することを狙ったシステムを構築した。

1 はじめに

GPTをはじめとした大規模言語モデル (LLM) の発展は著しく [1], 近年の LLM 利活用は、研究開発のみならず様々な事業・サービスへと広がりを見せている。LLM が高度な言語能力を持つことは広く知られており、従来困難とされていた自然言語処理タスクにおいて人間並みの精度を達成している [2]. 対話タスクも例外ではないが、雑談対話では既に一定の性能が確認されているのに対し、タスク指向対話での検証が十分とは言えず、様々な評価が試みられている [3, 4]. 対話ロボットコンペティション 2023 (DRC2023) も評価型ワークショップとして開催されており、旅行代理店における観光案内を模したタスクで、ヒューマノイドロボットを利用した高度なマルチモーダル対話が要求される [5].

我々は DRC2023 の観光案内タスクを対象として、人間のユーザの要望に応えながらインタラクティブに観光プランニングを行うシステムを提案する。本システムは LLM による生成を基本に柔軟な対話を行いながら、オープニング、ヒアリング、コース案内とスポット決定、旅行プランの提案と質問応答、

クロージングと 5つのフェイズを順に遷移することで、ユーザにあった旅行プランを提案するシステムである。各フェイズでは外部 API から取得した画像や地図情報、経路情報などを活用し、効果的にユーザに観光情報を提案できるように工夫する。

DRC2023 を通して本システムの対話能力と計画立案能力を検証し、高度なタスク指向対話における LLM 対話研究の現在地とその課題を明らかにする。本研究の提案システムは DRC2023 の予選 4 位、本戦で 1 位を獲得しており、発展途上ながら一定の成果があったことも合わせて報告する。

2 DRC2023 のタスク説明

DRC2023 は前回大会 (DRC2022 [6]) と同様にカウンターセールスによる観光案内タスクによって、対話システムを評価する。対話を行うブースでは、机を挟んで、ヒューマノイドロボットと対面する形でユーザが椅子に座り、音声対話を行う。対話終了後にアンケートによる評価を行う。机上には指向性マイクとディスプレイが、ロボットの背後にはユーザを写すカメラが設置されている。ディスプレイには観光地名、観光地画像、地図の表示が可能である。

なお、DRC2022 との違いは音声対話時間が 10 分間に伸びたこと、京都市内全体を対象として 1 日で 2 カ所の観光地を巡る旅行プランまでを立てるタスクになったこと¹⁾、これらの要件を満たすためにより複雑な対話が要求されている。

3 提案システム

我々が DRC2022 に参加したシステムでは、LLM を応答生成に活用するものの、ルールでシナリオを制御する対話システムを実装していた [7, 8]. しか

1) 前回大会は既に用意された 2つの観光地から 1カ所を推薦するタスクだった。



図1 提案システムのフェイズ遷移のフロー

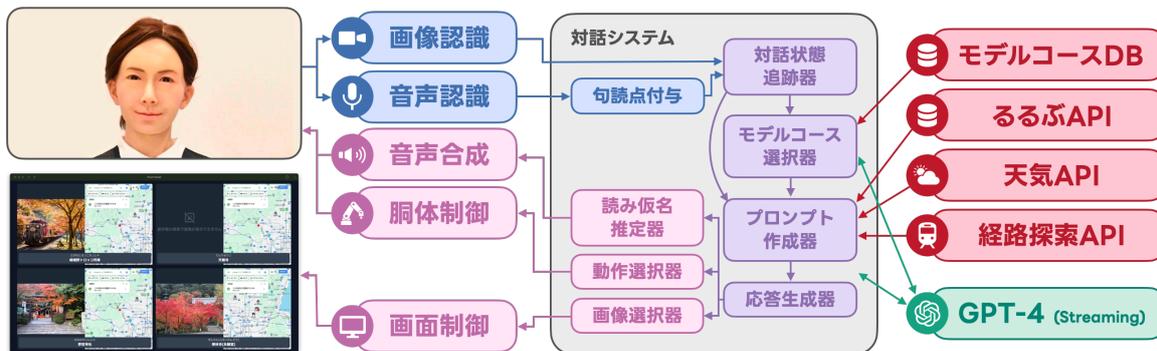


図2 提案システムの概要図。(左上画像はDRC2023公式サイトから引用。)

しながら、DRC2023ではより高度な対話が要求されることからルールベースでの制御は困難であると判断した。そこで本システムでは、LLMへの依存度を高め、応答生成を柔軟にするため、シナリオにおける各フェイズの粒度を拡大させた。さらには、観光地の増大やタスクの複雑さに対応するため、モデルコースデータセットや経路探索APIなどを使用し、知識ベースを拡大することで、検索拡張生成(RAG) [9]の活用を意識した設計となっている。

提案システムは事前に設定したシナリオ(図1)に基づいて、対話を進行する。シナリオは5つのフェイズで構成され、各フェイズそれぞれで対話内容が変わるように設計される。オープニングからスタートし、ユーザに対して簡単にヒアリングする。ヒアリング結果を元にモデルコースを提示し、ユーザと対話を行うことでモデルコースの決定し、その中から訪れる観光地2つを決定する。その後、旅行プランの提案や質疑応答を行い、最後に決まった観光地の確認と挨拶をして対話終了となる。

図2に提案システムの概要図を示す。提案システムは、マルチモーダル対話を実現するためにさまざまなモジュールから構成されている。ここではシステムの核となる機能とフェイズの一部を紹介する。

3.1 LLMによる応答生成

提案システムでは、ユーザの発話に対して柔軟な応答を返すために、フェイズのはじめなど一部の発

話を除き、LLMによって応答を生成する。応答の生成はプロンプトによって制御し、各フェイズではそれぞれ異なるプロンプトを使用する。

LLMにはOpenAI社より提供されている、gpt-4-32k-0613のAPIを使用する[10]。応答速度を向上のため、チャンクごとにレスポンスが得られるStreaming APIを使用し、句読点が表示される度に生成された内容を音声合成(TTS)に送信する。

3.2 フェイズ遷移

シナリオベースで進める際にはフェイズ間の遷移の制御が重要である。発話ターン数等でシナリオを遷移させると、ユーザが質問していてもそれを無視して、次のフェイズに進んでしまう可能性がある。

提案システムでは、フェイズを遷移させる方法として、フェイズ終了しても良いとLLMが判断した場合に、終了を表すトークンを出力させる。具体的には、各フェイズにおける目的が達成された時に“[END]”という終端記号を出力するように指示し、“[END]”が出力された時に次のフェイズに遷移する。これにより、例えばヒアリングフェイズにおいてユーザに聞くべきことを全て聞き終わった時に、自動で次のフェイズに遷移することができる。

しかしながら、LLMで終了を判定する手法では、すでにフェイズの目的が達成されているにも関わらず、“[END]”を出力しない場合がある。この問題に対処するため、フェイズごとに最大ターン数を設定

```

===
あなたの名前はショウコです... # 状況説明とペルソナの指示
2つのモデルコースを紹介し... # タスク指示

お客様に合ったモデルコースは以下の... # モデルコース情報
---
A コース: 紅葉が綺麗で... (スポット 1-スポット 2-...)
B コース: 有名な寺社仏閣が... (スポット 3-スポット 4-...)
=== # ショット
ショウコ: ...
お客様: ...
... # 最大で2ショット
=== # 現在進行中の対話
ショウコ: ... # 対話履歴
お客様: ...
...
ショウコ: # ここから生成開始

```

図3 スポット決定フェイズのプロンプト

し、そのターン数を超えた場合は強制的に次のフェイズに遷移するという制限を入れた。

3.3 カスタムコース生成とコース選択

本システムが想定するユーザは旅行代理店を訪問するお客様である。これらのユーザは独自に旅行プランを立てる人物像とは異なり、システム側からの提案を多く求める傾向にあると仮定した。そこで我々は、オープニングの雑談の情報と、最低限のヒアリング情報を元にユーザに合ったモデルコースを提案し、ユーザ側の意思決定の負荷を下げる戦略を採用した。

複数のサイトや書籍から全 24 件のモデルコースを収集した。それらのコースを元に旅行代理店のユーザを想定したペルソナと、それにカスタマイズしたコース（以下、カスタムコースとする）のペアを LLM で生成し、重複を排除して最終的に 13 のコースを保存した。

対話では、オープニングの雑談の情報とヒアリング情報を元に、13 のカスタムコースから LLM ベースのコース選択器が上位 2 つのコースを選択する。

3.4 コース案内とスポット決定フェイズ

ヒアリングによってカスタムコースが選択された後は、このコースの案内とスポットを決定する。本コンペティションのタスクとして、2 つの観光地を決定することは重要な要件であり、本フェーズは 5 つのフェーズでも重要な役割を持つ。

本フェーズでは提案システムの中核となるプロンプトを利用する。1 つのプロンプトにより、コースを紹介する対話を通して、間接的にユーザの好みをヒアリングし、2 つの観光地を決定するところまでを扱っている。

図 3 にプロンプトの概要を示す。プロンプトには前フェイズまでで決定された 2 つのカスタムコースの情報を外挿し、一度にまとめて情報を説明するのではなく、インタラクティブな対話の中で順番に観光地を紹介するよう指示を与える。また、常に観光地の画像を表示させながらユーザの反応を受け取り、コースの好みとその中で気に入ったスポットの決定までをリードして対話を行う。

3.5 外部情報を利用した応答生成

質問に対する応答や観光地間の経路を説明する際には、LLM で生成するだけでは、ハルシネーションを起こす可能性がある [11, 12]。そこで外部 API を利用して関連情報を使い、適切な応答を生成する。

二つの観光地を回る旅行プランを立てるには、最新の交通情報等も加味した正確な経路案内が必要となる。我々は NAVITIME API²⁾ を用いて取得した経路情報をプロンプトに追加した。ユーザの要望次第でランチ情報などの質問にも回答するために、プロンプトには観光地の基本情報もるるぶ API から取得して外挿する。また、オープニングでは気象庁の提供している天気 API³⁾ から得た情報を外挿しており、天気に関する雑談も可能である。

3.6 観光地画像の表示

音声テキストだけでは観光地の雰囲気が伝わらないため、観光地の画像を適切に表示することは、観光地を選択する上で重要な役割を果たすと考える。提案システムでは、観光地画像を表示するタイミングは 2 種類用意した。

一つ目は観光地名が LLM の発話内に登場した際にその画像を表示するというものである。同一発話内に複数スポットが登場した際は全てのスポットの画像を表示する。ただシステムの都合上 4 つより多く画像を表示できないため、5 つ目の画像以降は 4 つ目の画像を入れ替えることで対処した。

二つ目のタイミングとしては、モデルコースの名前が出た時にそのモデルコースに紐づけられた画像一覧を表示するというものである。この画像一覧はモデルコースを象徴するような画像が集まっており、ユーザに対して訴求効果が高いと考えられるため、一つ目の条件より優先して表示を行った。

2) <https://api-sdk.navitime.co.jp/api/>

3) <https://www.data.jma.go.jp/developer/index.html>

3.7 音声入出力の工夫

テキスト入力の場合と異なり音声入力では、句読点が付与されていなかったり、音声認識ミスによってユーザの発話と異なるものが入力されることがある。句点の有無が対話の応答生成に影響を与えるという報告もある [13]。そこで本研究では、言語モデルの生成確率を用いた句点の付与を行った [14]。

音声出力では、音声合成した際に読みを間違えて発音してしまう問題がある。京都の観光地名は特殊な読みをする場合もあり、他の一般的な単語でも読みに曖昧性があるものもある。この読みの問題に対しては、2つの方法で対処した。一つ目は、ルールベースの手法である。観光地名や今回のタスクにおいて読み方が限られるようなもの⁴⁾は、読み方の辞書をあらかじめ用意することで対処した。二つ目は、句点付与と同様に言語モデルの生成確率を応用して読みを推定した。複数読み方があって文脈に応じて読み方が変わるものをこの方法で対処する。

4 実験と結果

本節では DRC2023 の予選・本戦を通じた評価結果について述べる。

4.1 DRC2023 評価

評価は大きく分けて対話の自然さや満足度などを評価する対応評価と、旅行プランを立てるタスクを達成できたかを評価するプラン評価がある。予選会では 27 人の話者と対話し、予選順位 4 位で本戦に進出した。この 27 という有効対話数は全チーム中最大である。その理由として、全ての対話を生成で行うため、対話が容易には破綻せず対話を継続できるケースが相対的に多かったためと考えられる。

本戦では予選を勝ち抜いた 4 チームがそれぞれ 1 回ずつ対話を行う 1 次審査と、1 次審査を勝ち抜いた上位 2 チームで競う決勝が行われた。本システムは総合力で 1 次審査、決勝ともに 1 位の評価を受けて最優秀賞を獲得した。他チームとの相対比較では、本システムが最も LLM による生成を有効に活用し、頑健なシステムを構築していたことが高評価を受けた要因と言える。評価結果の詳細は Appendix A に参考として記載する。

4) 例えば、一日は「ついたち」と「いちにち」の 2 種類の発音で読めるが、本タスクでは「ついたち」と発音することはほぼない。

4.2 考察

次に本システムの戦略がどのように機能したのかを予選会中の具体的な対話例を通して考察する⁵⁾。まずアイスブレイクを利用して間接的にヒアリングを行う戦略では、過去の旅行経験の話をする中で、韓国旅行で食事を楽しんだことをユーザが話していたケースから、グルメを楽しめるコース提案に繋がった対話例などが存在した。

次にモデルコースを利用した段階的な絞り込みの戦略についても機能している事例が多く見られた。特に DRC2023 の設定上、ユーザは京都旅行に本当に行きたいわけではなく、「景色が綺麗ならどこでもいいですけど。」のような発話もあり、目的地の強い希望がないケースも多い。そのようなケースでも、モデルコースを 2 つ紹介し、選んでもらうことで効果的に対象範囲を狭め、さらにコース内の観光地を案内していくことで対話をリードして、スムーズな提案に繋がられている事が確認できた。

一方で対話中に起こるエラーとしては、音声認識誤りに起因するものが特に多く見られた。例えば「源光庵」のように「現行案」や「現行犯」など同音異義語や類似音が複数あるケースでは誤りが頻発し、これをきっかけに観光地を取り違えたり、存在しないモデルコースを案内してしまうケースも見られた。また、本システムで利用した特殊記号出力によるフェイズ遷移がうまく機能しないケースがあった。特にコース案内・スポット決定のフェイズに留まり続け、観光地を決定できないパターンが複数回確認された。その他、発話生成の待ち時間が長くなった結果、相手の話者と音声が被ってしまうなど、音声特有の問題も確認でき、ターンテイキング等の他分野と融合した、より高度なマルチモーダル課題に取り組む必要性も示唆された。

5 おわりに

本稿では DRC2023 に参加した音声対話システムについて、観光プランニングを行う上での様々な工夫について詳述し、コンペティションの評価を通して得られた知見を報告した。今後は確認されている問題点を改善するとともに、観光以外のタスクや他分野への応用を検討する。また LLM の利用についても、Instruction Tuning [15] や大規模化による LLM の性能と制御性を向上させる手法を模索したい。

5) 参考となる対話事例は Appendix B に記載する。

参考文献

- [1] Tom Brown, Benjamin Mann, Nick Ryder, et al. Language models are few-shot learners. In Advances in Neural Information Processing Systems, Vol. 33, pp. 1877–1901, 2020.
- [2] OpenAI, Josh Achiam, Steven Adler, et al. GPT-4 Technical Report, 2023.
- [3] Yi-Ting Yeh, Maxine Eskenazi, and Shikib Mehri. A Comprehensive Assessment of Dialog Evaluation Metrics. In Proceedings of The First Workshop on Evaluations and Assessments of Neural Conversation Systems, pp. 15–33, 2021.
- [4] Ryuichiro Higashinaka, Kotaro Funakoshi, Michimasa Inaba, et al. Dialogue System Live Competition: Identifying Problems with Dialogue Systems Through Live Event, pp. 185–199. Springer Singapore, Singapore, 2021.
- [5] Takashi Minato, Kurima Sakai Ryuichiro Higashinaka, Tomo Funayama, Hiromitsu Nishizaki, and Takayuki Nagai. Overview of dialogue robot competition 2023. In Proceedings of the Dialogue Robot Competition 2023, 2023.
- [6] Takashi Minato, Ryuichiro Higashinaka, Kurima Sakai, et al. Overview of Dialogue Robot Competition 2022. In Proceedings of the Dialogue Robot Competition 2022, 2022.
- [7] Takato Yamazaki, Katsumasa Yoshikawa, Toshiki Kawamoto, et al. Tourist Guidance Robot Based on HyperCLOVA. In Proceedings of the Dialogue Robot Competition 2022, 2022.
- [8] Takato Yamazaki, Katsumasa Yoshikawa, Toshiki Kawamoto, et al. Building a hospitable and reliable dialogue system for android robots: a scenario-based approach with large language models. Advanced Robotics, Vol. 37, No. 21, pp. 1364–1381, 2023.
- [9] Patrick Lewis, Ethan Perez, Aleksandra Piktus, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In Proceedings of the 34th International Conference on Neural Information Processing Systems, 2020.
- [10] OpenAI. Chatgpt, 2023. <https://chat.openai.com/>.
- [11] Ziwei Ji, Nayeon Lee, Rita Frieske, et al. Survey of Hallucination in Natural Language Generation. ACM Computing Surveys, 2022.
- [12] Vikas Raunak, Arul Menezes, and Marcin Junczys-Dowmunt. The Curious Case of Hallucinations in Neural Machine Translation. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1172–1183, 2021.
- [13] Tomoya Mizumoto, Takato Yamazaki, Katsumasa Yoshikawa, et al. Dialogue Systems Can Generate Appropriate Responses without the Use of Question Marks?—Investigation of the Effects of Question Marks on Dialogue Systems—, 2023.
- [14] 山崎天, 川本稔己, 大萩雅也ほか. Hyperclova を用いた音声雑談対話システム. 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, p. 22, 2022.
- [15] Long Ouyang, Jeffrey Wu, Xu Jiang, et al. Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, Vol. 35, pp. 27730–27744, 2022.

A DRC2023 評価結果の参考値

表1 予選評価結果（上位4チーム）

Team	有効数	対応評価平均	プラン評価平均
MIYAMA	17	5.817	0.941
ROS	20	5.150	0.850
JERRY cis	14	4.476	0.857
本システム	27	4.412	0.852

本戦は5人の審査員(A~E)で総合的に評価される。

表2 本戦1次審査結果
(A~Eは審査員毎の総合点)

Team	A	B	C	D	E	合計
本システム	17	21	15	16	13	82
ROS	12	19	14	19	16	80
JERRY cis	13	18	14	19	16	80
MIYAMA	11	21	20	11	12	75

表3 本戦決勝結果
(A~Eは審査員毎の総合点)

Team	A	B	C	D	E	合計
本システム	18	24	23	23	19	107
ROS	11	17	15	119	14	76

B 対話事例（参考）

ショウコ 旅行好きなんですね。一番印象に残っている旅行先や思い出、何かありますか？
 お客様 韓国旅行が思い出に残っています。
 ショウコ 韓国旅行ですか、素敵ですね！韓国と言えば美味しい料理や韓流ドラマ、K-POPが思い浮かびますね。お客様は何が一番印象に残りましたか？
 お客様 食べ物が美味しかったです。
 ショウコ 食事が美味しかったんですね！やっぱり旅行の色々な風景や体験も大切だけど、美味しいご飯も旅行の大きな楽しみですね。それでは、そんな食好きなお客様におすすめの京都のグルメスポットも提案させていただきます！
 ...
 ショウコ ...そして2つ目のコースは、「グルメと観光を楽しむコース」です。こちらは京都らしいお寺や神社だけでなく、美味しい食事やショッピングも楽しめる清水寺エリアや祇園・河原町エリアを巡るコースとなっています。

図4 アイスブレイクによる間接ヒアリング

ショウコ まずは1つ目のモデルコース、「家族で楽しむ京都コース」です。こちらでは、家族全員が楽しめるスポットを巡ります。嵐山の渡月橋や天龍寺から始まって...このコースなら、大人も子供も一緒に盛り上がること間違いなしですよ！
 お客様 これじゃない本がいいかな。
 ショウコ なるほど、そうでしたか。では、2つ目のモデルコースをご紹介します。これは「京都定番の神社仏閣巡りコース」です。こちらのコースでは、金閣寺や清水寺などの有名な寺院から、伏見稲荷大社や八坂神社など、見どころいっぱいのお寺を...
 お客様 すごくいいと思います。
 ショウコ 良かったです！では、「京都定番の神社仏閣巡りコース」で準備を進めてまいります。それぞれのスポットでどんな体験ができるか、詳しくはこれから一緒にプランを考えていきましょうね。モデルコースを選んでいただきましたので、このコースを元にお客様に合った観光地を2つ考えていきたいと思えます。よろしいでしょうか？
 ...
 ショウコ それではまずは鹿苑寺（金閣寺）から始めてみてはいかがでしょうか。きらめく金色の美しい建物とその周囲の庭園は、大人から子どもまで楽しめますよ。また、御朱印帳に金閣寺の御朱印をいただいでみるのも良い記念になると思いますよ。なお、「京都定番の神社仏閣巡りコース」では、八坂神社や伏見稲荷大社など、ほかにもたくさんのお寺があります。どこを訪れてみたいですか？
 お客様 景色がいいお寺だった。どこでもいいですけど。
 ショウコ 景色がいいお寺といえば、「常寂光寺」はいかがでしょうか？ここは紅葉のすばらしいことで有名なお寺で、特に秋に訪れると、庭園が美しく染まるんですよ。四季折々の風情が楽しめるので、とてもおすすめです。さて、そんな常寂光寺と鹿苑寺（金閣寺）、この2つを訪れるプランはいかがでしょうか？

図5 モデルコースを活用した提案例