

疑似対話データと Preference データを用いた ドメイン特化対話への日本語 LLM チューニングの検証

山崎 天¹ 高山 隼矢¹ 佐藤 京也^{1,2} 大萩 雅也¹ 吉川 克正¹ 水本 智也¹

¹SB Intuitions 株式会社 ² 東京都立大学

{takato.yamazaki, junya.takayama}@sbintuitions.co.jp

概要

本研究は、日本語で学習された大規模言語モデル (LLM) を観光対話システムに適応させるため、効率的なデータ作成とチューニング手法を検証することを目的とする。疑似対話データを用いた Supervised Fine-Tuning (SFT) および少量の Preference データを用いたチューニングを実施し、観光案内タスクにおける対話の自然性、一貫性、満足度、信頼性への寄与を評価した。実験では、汎用指示チューニングのみのモデル、SFT モデル、Preference モデルを比較した結果、SFT モデルは対話の自然性と満足度を大幅に向上させる効果が確認された。一方、Preference モデルでは Hallucination の抑制が可能であったが、満足度の向上には課題が残った。

1 はじめに

近年、大規模言語モデル (LLM) は、汎用的な指示データを用いた指示チューニング (Instruction-Tuning) によって高い性能を発揮することが広く認識されている [1]。しかし、汎用的な指示チューニングだけでは、観光案内や施設ガイドのような特定のドメインにおいて理想的な生成結果が得られない場合があり、ドメイン特化が求められるケースも少なくない。こうしたドメイン特化対話システムの開発には、以下のような課題が存在する。第一に、新たなドメインに適応させるための十分な学習データを用意するには、多大な労力とコストがかかる。第二に、対話の一貫性を保ちつつ、ドメイン外の応答や誤情報 (いわゆる Hallucination) を抑制する必要がある。これらの課題を解決するには、効率的かつ効果的な LLM のチューニング手法が求められている。

本研究では、LLM のドメイン特化に向けた効率的なデータ作成とチューニング手法を検証することを目的とする。具体的には、日本語で学習された LLM に対し、対象ドメインの疑似対話データを

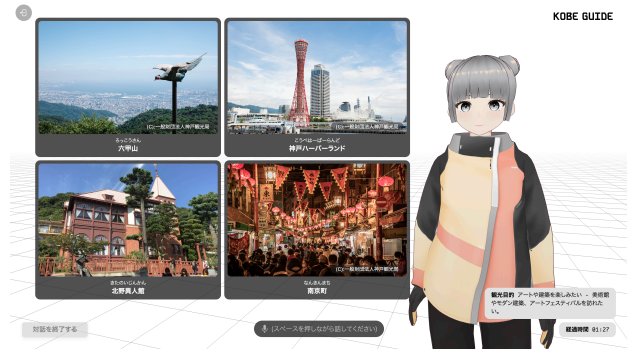


図 1 本研究で実験対象とした観光案内ドメインのAvatar音声対話インターフェース。左側には対話中に言及された観光地の画像が表示される。

用いた教師ありファインチューニング (Supervised Fine-Tuning, SFT) [2, 3]、および人手で作成された少量の Preference データを用いたチューニング [4, 5, 6] を行う。さらに、Preference チューニングでは、知識やデータの不足による Hallucination を起こしている発話に対し、回答拒否を行うように人手で修正した発話を作成し、そちらを選好されたデータとして訓練した。実験では観光案内ドメインを対象とし、汎用ドメインの指示チューニングのみのモデル (Instruction-Only)、疑似対話データを用いた SFT モデル (SFT)、SFT モデルに Preference チューニングを施したモデル (Preference) の 3 種類の LLM を搭載したAvatar対話システムを実装した。これらのモデルを、人手によるアンケート評価を通じて、対話の自然さ、満足度、信頼性、そしてドメインに対する一貫性の観点から比較した。結果として、疑似対話データによる SFT モデルは汎用的な指示チューニングのみのベースモデルと比較して大幅に対話の自然性や満足度が向上した。一方で、Preference モデルは情報不足の質問に対して適切に回答拒否を行う能力を示したものの、ユーザー満足度は低下する結果となった。これにより、Hallucination を抑制することが必ずしもユーザー満足度の向上につながらないことが明らかになった。

本研究の貢献点は、ドメイン特化型対話システムにおけるチューニング手法を包括的に検証し、それぞれの効果と課題を明らかにした点である。これにより、今後のシステム設計や開発プロセスで考慮すべきポイントを示し、ドメイン特化の際に役立つ知見を提供することを目指した。

2 実験設定

観光対話タスク 本研究で対象とする観光対話タスクは、対話システムロボットコンペティション 2024[7, 8] のタスクをもとに設計したもので、観光目的を持ったユーザーと 3D アバターによる音声対話を通じて観光地の選定や提案を行うものである。ユーザーの観光目的の例としては、「温泉で癒やしを求めたい」や「写真映えスポットを巡りたい」のようなものが挙げられる。対話システムには事前に観光エリア（本実験では神戸）とその周辺の観光地情報が与えられており、ユーザーの発話内容に基づき、観光目的に合致した場所を提案し、それに関する情報を提供する。対話の時間は約 3 分間とした。本研究での実験では、被験者に多様な観光目的で対話を実施してもらうため、対話開始前に事前に ChatGPT[9] を用いて作成した 40 個の観光目的をランダムに割り当て、その目的を満たしたい観光客としてロールプレイを実施させた。

対話インターフェース 被験者が使用する対話インターフェースを図 1 に示す。Web 上に実装されており、ユーザーはスペースキーを押しながら音声入力を行う。アバターの左側には、システムが発話の中で言及した観光地の画像が表示される。アバターの動作は、発話時にあらかじめ選定された動作からランダムに割り当てたものと、観光地画像の表示時に視線を誘導する動作の 2 種類を実装した。

対話システム 対話システムについては、音声認識器に内蔵される音声区間検出器（Voice Activity Detection）がユーザー発話の終了を検知した際、その内容を LLM に送信し、応答を生成するようなシステム構成とした。本実験ではこの LLM のチューニングを通じて観光ドメインにおける対話性能の改善を目的とする。

3 SFT 用疑似対話データの作成

LLM を単一ドメイン向けにチューニングするために、人手でデータを作成するには多大な費用と労力がかかる。そこで、本研究では LLM を利用して

疑似対話データを生成し、指示チューニング済みの LLM に対して追加で疑似データによる教師ありファインチューニング (Supervised Fine-Tuning, SFT) を行う。具体的な疑似対話データの生成プロセスは以下の通りである。

1. **観光データの収集:** Web サイトからクロールリングを用いて観光データを収集した。このデータは観光地のタイトルと概要情報で構成される。
2. **ランダムサンプリング:** クロールリングで得られた観光地データの中から対話中に使用するものをランダムに 5 ~ 10 件を選択した。これにより、対話内容に多様性を持たせる。
3. **対話スクリプトの生成:** 選択した観光データを含む、対話スクリプト生成のためのプロンプトを設計した (付録 A の図 4 を参照)。プロンプトを指示チューニング済みの LLM である Nemotron-4-340B-Instruct[10] に入力し、疑似対話データを生成した。

本研究では、観光地データは神戸、香川、熊本、東京の 4 種類のエリアから収集し、12 ターン程度で構成される対話データが各エリアについて 3,000 件、合計 12,000 件の疑似対話データを作成し、訓練に使用した。作成されたデータの例は付録 A の表 3 に示した。その後、生成された疑似対話データを用い、LLM の Fine-Tuning を実施した。

4 Preference チューニング

疑似データを用いてチューニングしたモデルをさらに人間の選好に近づけるため、Preference データを作成し、チューニングを行った。

Preference チューニング用データは、4 人のアノテーターによって作成した。アノテーターには、ユーザー発話に対する 2 種類の応答を LLM で生成し、それらを比較評価するための専用 UI を提供した。アノテーターは、どちらの応答がユーザーにとっての理想に近いかを判断し、評価を 7 段階で行った（「A(B) が明確に優れている」、「A(B) が優れている」、「A(B) が少し優れている」、「等しい」）。各対話は 5 ~ 10 往復の発話を目安とし、合計 48 対話を収集した。

応答の修正 LLM が生成する応答の品質が低い場合には、アノテーターが手作業で理想的な応答へ修正または新たに執筆する工程を組み込んだ。修正・執筆された応答は最高品質のものとみなし、LLM が生成した応答とペアにして選好データとし

て訓練データに加えた。

さらに、アノテーターにはシステムが使用している観光データを提示し、対話中に観光データに含まれない情報(例: 入場料や電話番号など)について積極的に質問させ、モデルがこれらに誤って答えた場合には、適切に拒否する応答(例: 「申し訳ありません。データベース内に情報がありません。お答えできません。他に行きたい場所がありますか?」)を記述するよう指示した。これにより、Hallucinationの低減を目指した。

SimPO による訓練 収集した48件の対話データから305組のペアデータを生成し、擬似対話でSFTしたモデルに対して追加の訓練データとして活用した。PreferenceチューニングのアルゴリズムにはSimPO (Simple Preference Optimization)[6]を採用した。SimPOは、従来の手法であるDPO (Direct Preference Optimization) [5]と比較して計算負荷が低く、高速な最適化を実現するだけでなく、性能面でも優位性を示すアルゴリズムである。

5 実験

5.1 評価方法

§2で説明したインターフェースを使用し、アノテーションの専門家10名に各モデルとの対話とその評価を依頼した。

比較モデル 以下の3種類のモデルを比較した。

- **Instruction-Only:** Sarashina2-70B[11]に汎用ドメインの指示チューニングを施したモデル
- **SFT:** Instruction-Onlyモデルをベースに観光分野に特化した擬似対話データを用いてSupervised Fine-Tuningを施したモデル
- **Preference:** SFTモデルをベースにPreferenceデータでSimPOを追加適用したモデル

モデルに与える指示文や観光データは統一した。

評価手順 被験者には2つのモデルを対象に、3分間の対話とアンケート評価を行わせた後、2つのモデルを比較評価させた。これを1ラウンドとし、各被験者には7~8ラウンドを実施してもらい、合計75ラウンド分の評価データを収集した。各ラウンドではモデルの順番をランダムに設定し、被験者にはモデル名を伏せた。また、同じラウンド内では同じ観光目的で対話を行うよう指示した。

各モデルとの対話後のアンケートでは、満足度やドメイン適応性を評価するため、以下の4つの評価



図2 各モデルのアンケート評価の結果

項目を設定し、それぞれ4段階で回答させた。

- 自然な対話ができましたか? (Naturalness)
- 満足に観光地について知りたい情報を得られましたか? (Satisfaction)
- 提供された情報は信頼できると感じられますか? (Trustworthiness)
- 一貫して観光客の目的に沿った提案や回答ができていましたか? (Consistency)

比較評価においては、「どちらがより優れた音声案内ができていましたか?」という質問に対して5段階評価(「A(B)がとても優れている」、「A(B)が優れている」、「等しい」)で回答させた。被験者が対話内容を忘れることを防ぐため、比較評価中には各システムとの対話ログを表示した。

5.2 結果

各モデルに対するアンケート評価の結果を図2に示す。SFTは4つの評価指標すべてにおいて他モデルを上回り、特にNaturalnessとConsistencyにおいて顕著な差を示した。Instruction-Onlyは全体的に最も低い評価を受けたことから、擬似対話データによるSFTはユーザー評価の向上に有効であり、満足度が高まる応答が生成できるようになることがわかった。一方で、PreferenceはSFTを下回る結果となったが、これは我々の予測に反する結果であった。すべての評価指標でInstruction-Onlyを上回ったものの、SFTには及ばず、改善の余地があることがわかった。

図3にモデル間の比較の結果を示す。比較評価においても、SFTとPreferenceの両モデルがInstruction-Onlyモデルより大幅に優れていることが確認できる。SFTとPreferenceを比較した場合、SFTの方が全般的に優れた評価を得る結果となった。

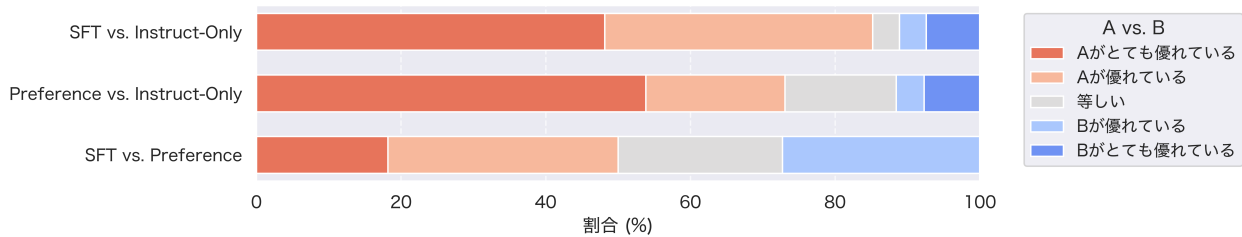


図3 モデル間の比較評価の結果。

5.3 考察

Hallucination 回避のための回答拒否の影響 表1に示す SFT の対話例では、下線部で Hallucination を発生させており、観光案内としては問題であるが、回答できたという点で対話の満足度は上がってしまった可能性がある。一方で、表2の Preference の例では、与えられた観光データでは回答できない質問に対し、適切に拒否しており、Hallucination を低減させるという我々の目的は反映できたといえる。

しかし、「申し訳ありません」という発言が含まれる発話は全発話の17%を占めており、その高頻度な謝罪がユーザー満足度の低下や不信感の増加を招いたと考えられる。原因として観光データの不足が挙げられるため、モデルに十分な情報を与えることでユーザー満足度の向上が期待できる。一方で、「申し訳ありません。神戸では、雪まつりのようなイベントは行われておりません。」といった、適切な情報提供ができていながらもかわらず、不必要に謝罪する例も散見された。これは、不十分な量の Preference データによって過学習が生じた可能性があり、その改善は今後の課題となる。

User	寒い冬でも楽しめるお祭りって何かないかな
Assistant	もちろん、ありますよ。神戸では毎年1月に「神戸ルミナリエ」というイルミネーションイベントが開催されています。震災の記憶を後世に語り継ぎ、神戸の夢と希望を象徴する行事として定着しています。
User	今年はいつからやってるの
Assistant	今年の開催期間は12月6日から12月15日までです。もし期間中に神戸にお越しの際は、ぜひお立ち寄りください。
User	結構人が来て混むんだよね 大丈夫かな
Assistant	はい、例年多くの人で賑わいます。もし混雑が気になるようでしたら、平日の夜にお越しいただくことをおすすめします。
User	電車で行くとしたらどこの駅で降りれば良い
Assistant	最寄り駅はJR元町駅で、駅から徒歩約10分ほどで会場に到着します。

表1 SFT モデルの対話例。下線部では Hallucination が発生している。

生成文の長さの影響 各モデルの発話の平均文字数は、Instruction-Only が328.8文字、SFT は56.6

User	有馬温泉の営業時間って何時から何時
Assistant	申し訳ありません。データベース内に情報がありません。他に知りたいことはありますか？
User	有馬温泉の電話番号ってわかる
Assistant	申し訳ありません。データベース内に情報がありません。
User	有馬温泉って何種類くらいの温泉があるの
Assistant	有馬温泉は金泉と呼ばれる赤褐色の含鉄泉が有名です。また、透明な銀泉も湧出しています。

表2 Preference モデルの対話例。

文字、SimPO は61.9文字であり、Instruction-Only はSFT の約6倍もの文長であった。この違いは、モデルの訓練に使用された汎用的な指示データが書き言葉を前提としており、長文の応答が多く含まれていることに起因すると考えられる。Instruction-Only は観光情報を包括的に説明する傾向があり、ユーザーの質問に対して冗長な応答を生成する場合が多かった（対話例は付録Bの表4を参照）。指示プロンプトには「短く簡潔に出力する」よう明記していたが、その効果は限定的であった。

一方、表1に示す SFT の対話例では、疑似対話データに基づく調整により、発話の文長が音声対話に適した水準に収まっていた。この結果、ユーザーとの対話で質問や趣向を引き出す頻度が増え、満足度の向上に繋がったと考えられる。

6 おわりに

本研究では、日本語で学習された大規模言語モデル (LLM) を観光ドメインに適応させるため、疑似対話データを用いた Supervised Fine-Tuning (SFT) と、少量の Preference データを活用したチューニングの効果を検証した。その結果、SFT は対話の自然性およびユーザー満足度の向上に有効であることが示された。一方で、Preference チューニングは Hallucination の抑制に一定の効果を発揮したものの、ユーザー満足度の向上には寄与しなかった。今後の課題として、データ作成プロセスの改善などを通じて、ユーザー評価を維持しながら Hallucination のさらなる抑制を図る必要がある。

参考文献

- [1] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. **Advances in neural information processing systems**, Vol. 35, pp. 27730–27744, 2022.
- [2] Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A Smith, Daniel Khashabi, and Hannaneh Hajishirzi. Self-instruct: Aligning language models with self-generated instructions. In **The 61st Annual Meeting Of The Association For Computational Linguistics**, 2023.
- [3] Zhangchen Xu, Fengqing Jiang, Luyao Niu, Yuntian Deng, Radha Poovendran, Yejin Choi, and Bill Yuchen Lin. Magpie: Alignment data synthesis from scratch by prompting aligned llms with nothing. **arXiv preprint arXiv:2406.08464**, 2024.
- [4] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In **Advances in Neural Information Processing Systems (NeurIPS)**, 2017.
- [5] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. **Advances in Neural Information Processing Systems**, Vol. 36, , 2024.
- [6] Yu Meng, Mengzhou Xia, and Danqi Chen. Simpo: Simple preference optimization with a reference-free reward. **arXiv preprint arXiv:2405.14734**, 2024.
- [7] Takashi Minato, Ryuichiro Higashinaka, Kurima Sakai, Tomo Funayama, Hiromitsu Nishizaki, and Takayuki Naga. Overview of dialogue robot competition 2023. **arXiv preprint arXiv:2401.03547**, 2024.
- [8] Takato Yamazaki, Katsumasa Yoshikawa, Toshiki Kawamoto, Tomoya Mizumoto, Masaya Ohagi, and Toshihiko Sato. 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] OpenAI. ChatGPT. <https://chatgpt.com>. Accessed: 2025-01-08.
- [10] Bo Adler, Niket Agarwal, Ashwath Aithal, Dong H Anh, Pallab Bhattacharya, Annika Brundyn, Jared Casper, Bryan Catanzaro, Sharon Clay, Jonathan Cohen, et al. Nemotron-4 340b technical report. **arXiv preprint arXiv:2406.11704**, 2024.
- [11] SB Intuitions, Inc. 日本語性能トップクラスの llm・sarashina2-70b の公開. <https://www.sbintuitions.co.jp/blog/entry/2024/08/21/144254>, aug 2024. Accessed: 2025-01-08.

A 疑似対話の作成

図 4 には疑似対話の作成時に LLM に与えたプロンプトを示す。表 3 には実際に生成された疑似対話データの例を示す。

Use the following information to generate a dialogue script of a guide ナオ and a customer who is planning to go to {{ area }} for sightseeing.
ナオ starts by greeting and gently asking about their sightseeing and preferences. Keep responses friendly, brief, and conversational.
Use the Tourist-Spot Information during the guide.

ナオ's Persona
- Speaks Japanese fluently
- Polite and Friendly
- A woman with silver hair

Example Output
The following is an example output.
...
ナオ: こんにちは。ようこそ神戸へ。神戸は楽しんでいますか？
お客様: はい、楽しんでいます。昨日は三宮で明石焼きっていうんですか？美味しかったです。
ナオ: 明石焼きいいですね！その他にも色々美味しいものがありますが、ご興味ありますか？
お客様: おすすめはありますか？
ナオ: 神戸ビーフはやっぱ有名ですが、結構いいお値段がするので、私のおすすめとしては神戸ラーメンやそばメンなどですね。
...
Tourist-Spot Information
Use some of the following data in the dialogue script.
{{ sight_data }} # 形式は「観光地名: 観光地概要」

Direction
Now, create a dialogue script between ナオ and the customer about sightseeing in {{ area }}, using the Tourist-Spot Information. Begin ナオ's lines with 'ナオ: ' and the customer's lines with 'お客様: '. Start the conversation with a casual introduction about tourism. Encourage the customer to ask detailed questions about the recommended tourist spots, even beyond what can be answered solely with the Tourist-Spot Information. If the Tourist-Spot Information is insufficient to answer a question, ナオ can provide a response based on her personal knowledge. However, ナオ should only answer questions if she is confident in her knowledge. Generate approximately 12 exchanges in the dialogue.

図 4 疑似対話の作成に使用したプロンプト。Nemotron の学習データの主要言語に合わせ、英語で記述した。

B 対話例

表 4 には Instruction-Only モデルの対話例を示す。

Assistant	こんにちは！熊本へのご旅行、楽しみです。熊本にはたくさんの魅力的な観光スポットがありますよ。
User	こんにちは。熊本ラーメンが有名ですね。おすすめのラーメン屋さんを教えてください。
Assistant	熊本ラーメンといえば、黒亭と龍虎ラーメンが有名です。黒亭は老舗で、細くて真っすぐな麺にとんこつスープが特徴です。龍虎ラーメンは創業 47 年の老舗で、こってりな豚骨スープと甘めのチャーシューが人気です。
User	どちらも美味しそうですね。歴史的な観光スポットも興味があります。
Assistant	でしたら、立田自然公園はいかがでしょう。細川家の菩提寺泰勝寺の跡で、京風の雅を残す庭園があります。また、加藤清正銅像もおすすめです。
User	加藤清正銅像はどこにあるんですか？
Assistant	熊本城の近くにあります。熊本城も一緒に観光するといいですね。
User	温泉も楽しみたいのですが、おすすめはありますか？
Assistant	植木温泉はいかがでしょう。源泉かけ流しの豊富な湯量があり、泉質は弱アルカリ単純イオン泉で、湯上がりの肌はしっとりします。
User	温泉もいいですね。ショッピングも楽しみたいのですが、おすすめの場所はありますか？
Assistant	サクラマチクマモトはいかがでしょう。熊本城と庭づき、すべてをつなぐ「おもてなしの庭」で、ファッション・雑貨・コスメ・飲食など 149 店が集まる新しいライフスタイル発信地です。
User	楽しそうですね。熊本城はどんな歴史があるんですか？
Assistant	熊本城は、戦国時代末期から江戸時代初期にかけて活躍した武将、加藤清正が築城した城です。明治時代には西南戦争の舞台となり、現在は国宝に指定されています。
User	興味深いですね。熊本城の近くに美味しいお店はありますか？
Assistant	熊本城の近くには、美味しい熊本ラーメンのお店や、熊本名物の馬刺しが食べられるお店がありますよ。
User	ありがとうございます。たくさん観光できそうですね。
Assistant	ぜひ、熊本を満喫してくださいね。何かあれば、またお尋ねください。
User	ありがとうございます。楽しみです。
Assistant	では、よい旅を！

表 3 Nemotron によって生成された疑似対話の例。

Assistant	こんにちは。神戸の観光案内担当の「ナオ」です。よろしくお願いします。
User	よろしく
Assistant	お客様が神戸の観光や食事で好まれるものについて、少しお伺いしてもよろしいでしょうか？例えば、自然や歴史的な名所がお好きなのか、それとも現代的な建物やショッピングがお好きなのか、また、食事では神戸牛のような高級グルメが良いのか、...
User	食事はひとりで静かに食べたいんだけどなんかある
Assistant	おひとりで静かに食事をされたいのですね。神戸には、ひとりでも入りやすいお店や、落ち着いた雰囲気のお店がたくさんあります。例えば、三宮駅周辺には、ひとりでも気軽に入れるラーメン店や、定食屋、カフェなどがたくさんあります。特に、三宮センタープラザやさんプラザの地下には、様々な飲食店が集まっております、...

表 4 Instruction-Only モデルの対話例。