

レコメンデーションにおける体験価値に基づいた提案文の生成

吉見菜那 高橋ゆり子

NTTドコモ サービスイノベーション部

nana.yoshimi.tw@nttdocomo.com yuriko.takahashi.rz@nttdocomo.com

概要

LLMを用いたレコメンデーション研究において、行動履歴や属性に基づく推薦理由の提示が試みられているが、商品の特徴を中心とした従来の説明では、ユーザーが具体的な利用シーンを想起することは困難である。本研究では、ユーザーのインサイトを深層的に分析し、パーソナルデータから推測されるライフスタイルに即した体験イメージを物語形式で提示する手法を提案する。本アプローチにより、単なる情報提示を超えた深い納得感を醸成し、より満足度の高い意思決定および購買体験の実現をめざす。

1 はじめに

近年、大規模言語モデル (LLM) の台頭によって高度な論理的思考と文章生成が可能となり、広範なタスクにおいて革新的な成果が得られている。商品やサービスをユーザーに推薦するレコメンデーションの分野においても、その進展は顕著である。従来の協調フィルタリング[1]等に代わる新たなアプローチとして、LLMの活用が急速に進んでおり、双方向の対話を通じたユーザー嗜好の把握や自然言語による柔軟な表現力は、従来の推薦システムにはない独自の強みとなっている。

こうしたパーソナライズの高度化に伴い、レコメンデーションにおけるテーマの一つとして、提案内容の説明性に関する研究が取り組まれている[2-5]。LLMは、従来ブラックボックスであった商品の推薦根拠を文章として提示することを可能とし、ユーザーの納得感や推薦システムへの信頼性の向上に貢献している。しかし、既存研究の多くは主にユーザーの行動履歴や嗜好との整合性を語る説明に焦点を当てており、その商品が各ユーザーの日常生活でどのように役立つかという具体的な利用イメージが欠如しやすい。このような提案は、商品購入前の期待と購入後の実体験との乖離を招き、ユーザーの購買意欲の減退や購入後の満足度の低下を引き起こす要因

となり得る。

本研究では、ユーザーの潜在的な要望を深層的に分析し、個々のライフスタイルに即したメリットを物語形式で提示する手法を構築する。これにより、従来の嗜好適合性に基づく説明だけでは想起困難であった具体的な商品の体験のイメージを補完し、ユーザーの深い納得感と円滑な意思決定を支援することをめざす。

2 関連研究

2.1 LLMを用いた推薦手法

レコメンデーション分野におけるLLMの活用は、ユーザー個人にパーソナライズされた最適な商品を推薦する予測精度の向上に大きく寄与している。そのアプローチは、主に推薦指向、対話指向、シミュレーション指向の3つに大別される[6]。推薦指向のアプローチ[7,8]では、LLMの高度な推論能力によってユーザーの過去の行動履歴に含まれる意味的文脈を理解し、LLMを用いて直接推薦アイテムを選定することに特化している。対話指向のアプローチ[9,10]では、ユーザーとの双方向の対話を通じて意図や嗜好を汲み取り、文脈の遷移を考慮した自然かつ柔軟な提案文の生成を実現している。さらに、シミュレーション指向のアプローチ[11-13]では、ユーザーの行動履歴や嗜好を反映した仮想エージェントを構築し、実際のユーザーに代わって擬似的に推薦システムを評価させることで、従来の人手評価で課題となっていた人的、時間的コストを排除しつつ、大規模なインタラクション評価の再現を可能にしている。

2.1 LLMによる提案文の説明性

前述した対話指向のアプローチにおいて、LLMによる柔軟な文章表現が可能となったことを背景に、商品の推薦根拠の説明性に関する研究が取り組まれている。Gaoら[2]は、対話型フレームワーク

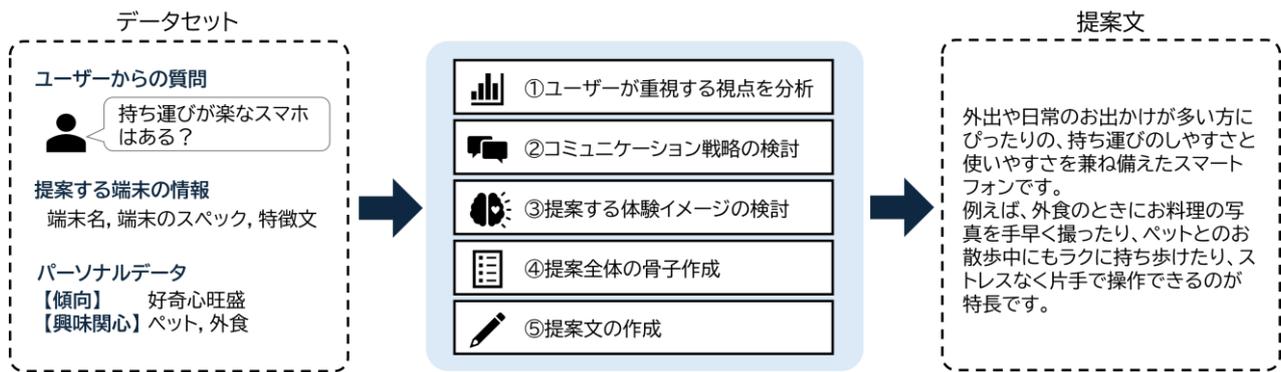


図 1 提案手法の概要

Chat-REC を提案し、プロンプトへの情報統合による説明生成の有用性を提示した。この流れを受け、ユーザーの過去の行動履歴や嗜好をプロンプトに明示的に記述し、個人の文脈に即した最適な説明を生成する手法が提案されている[3-5]。これらは、汎用的なテンプレートとは異なり、入力されたユーザー情報に基づいて動的に商品特徴を抽出し、推薦の根拠として提示する。個人の嗜好や文脈を反映させた説明は、画一的な情報提示と比較して高い受容性を示し、ユーザーの意思決定を促進することが実証されている。

3 提案手法

本研究では、商品特徴の羅列による体験イメージの欠如および期待と体験の乖離という課題を解決するため、ユーザーの潜在的な要望を分析し、推測されるライフスタイルに即したメリットを物語形式で提示する手法を構築する。提案手法の概要を図1に示す。本システムは、あらかじめ設定された推薦商品およびユーザーのパーソナルデータ（傾向、興味関心）を入力とし、LLM の CoT (Chain of Thought) プロンプティング[14]により以下 5 つのステップを段階的に実行して、推薦商品に対する提案文を作成する。

ユーザーが重視する視点的分析 ユーザーからの質問とパーソナルデータに基づき、ユーザーのプロファイルを整理する。また、対話から取得できる顕在的な要望だけでなく、行動の背景にある価値観や潜在的な要望を深層的に分析し、ユーザーが真に価値を感じるポイントを特定する。

コミュニケーション戦略の検討 パーソナルデータから得られるユーザーの傾向から、より納得感のある文章表現のアプローチを決定する。例えば、慎重な傾向のあるユーザーには客観的な事実や実績、

他ユーザーの評価に基づいて安心感を醸成するなど、ユーザーにパーソナライズされたコミュニケーション戦略を検討する。

提案する体験イメージの検討 商品の機能スペックとユーザーが重視する視点を照合し、“いつ、どこで、どのようにその商品がユーザーの生活を豊かにするか”という具体的な利用シーンを創出する。ここで、単なる機能説明を、ユーザーのパーソナルデータに基づく具体的な体験イメージへ変換する。

提案全体の骨子作成 提案内容と体験イメージの整合性のある提案骨子を設計する。具体的には、まずユーザーが重視する視点到訴するコアメッセージを定義し、商品特徴からそれがもたらす体験イメージを結びつける。さらに、ユーザーが自分事として没入できる物語形式の文章を描写してコミュニケーション戦略に基づいた骨子を作成する。

提案文の作成 作成された骨子に基づき、ユーザーに提示する提案文を作成する。

4 評価実験

4.1 実験設定

データセット 本実験では、図1に示すような入力データを使用する。まず、200件の仮想パーソナルデータを作成し、各パーソナルデータには傾向および興味関心をランダムに付与した。次に、10パターンのユーザーからの質問と提案端末のペアを用意した。質問には“充電を気にせず一日中使えるスマホが欲しい”といった具体的な要望が記述され、提案端末にはその内容に対応した端末名、端末スペック、特徴文が含まれる。実験では、200件の仮想パーソナルデータに対し、これら10パターンの質問と提案端末のペアをランダムに割り当てることで、パ

表 1 評価実験結果

カテゴリ	評価指標	自動評価における各手法の選択率 (%)			人手評価平均スコア	
		ベースライン	提案手法	引き分け	ベースライン	提案手法
情緒性価値	有効性	5.90	93.90	0.20	--	--
	説得力	1.50	98.50	0.00	--	--
	満足度	2.50	97.40	0.10	--	--
	共感性	2.30	97.70	0.00	--	--
	発見性	1.90	98.10	0.00	--	--
説明性評価	透明性	8.00	91.60	0.40	6.14	6.51
	信頼性	8.60	89.50	1.90	5.38	6.38
	精査可能性	10.10	88.30	1.60	5.34	5.99
	効率性	71.40	28.10	0.50	4.90	5.55
	視覚化可能性	3.20	96.50	0.30	3.62	4.50

ーナライズ化と課題解決を両立させる推薦タスクを構成した。

モデル 提案文の生成モデルには、Microsoft Azure OpenAI Service 上で提供される gpt-4.1¹⁾を使用した。

比較対象 提案手法の思考過程の有効性を検証するため、最終工程である“⑤提案文の作成”のみを実行する CoT を用いない手法をベースラインとして設定した。なお、入力データおよび実行環境は提案手法・ベースラインともに同条件である。

自動評価 生成された提案文の品質を LLM-as-a-Judge によって多角的に評価した。プロンプトにユーザーのパーソナルデータを含めることで、対象となるユーザーの視点をシミュレートさせた。評価モデルは、ベースラインと提案手法の出力を比較し、各評価指標においてどちらの手法が優れているかを選択させる二者択一形式とした。なお、公平性を保つため、プロンプトにはベースラインと提案手法がどちらかを推測させるような指示は行っていない。ここで、評価指標は推薦システムにおける説明の品質評価指標である Balog と Radlinski のフレームワーク[15]を基礎としている。本実験では、この既存指標に加え、体験価値に基づいた提案がユーザーに与える影響を測るための独自項目（共感性、発見性、視覚化可能性）を追加し、情緒性評価と説明性評価の2つのカテゴリからなる計10項目の評価基準を設定した。各項目の説明は付録に記載する。

人手評価 提案された文章が、人間にとって論理

的に整合し、かつ信頼に足る情報として機能しているかを検証するため、2名のアノテーターによる評価実験を実施した。なお、アノテーターに対しては、互いの評価は分からない状態で、提案手法とベースラインのどちらによる提案文かは伏せ、提案手法とベースラインの順序をランダムに混ぜた状態で評価を実施した。本評価では、第三者視点での客観的な妥当性を重視し、評価対象を説明性評価に絞って実施した。これにより、個人の嗜好に依存せず、情報の構造や透明性がどの程度担保されているかを測定する。なお、採点には7段階リッカート尺度を用い、提案の品質を絶対評価として判定した。

4.2 実験結果

自動評価および人手評価の結果を表1に示す。自動評価において、提案手法の選択率は情緒性価値の全5項目で93%を超え、一貫してベースラインよりも高い性能を示した。また、説明性評価においても5項目中4項目でベースラインの選択率を上回り、特に視覚化可能性では96.5%という高い値を記録した。一方、効率性においてはベースラインが71.4%と優勢であった。これはLLMが、読了時間が短い文章を効率性が高い文章として認識しており、体験イメージを充実させた提案手法の方が読了時間がかかるため、効率性が低く判定されたことが推察される。人手評価においては、評価値の平均値を比較すると提案手法は効率性を含む全ての項目でベースラインを上回る結果となり、本手法の有効性を確認できた。

表2に、各手法における提案文の出力例の一部を示す。本提案文は入力としてショッピング、旅行に

¹⁾ <https://azure.microsoft.com/ja-jp/pricing/details/cognitive-services/openai-service/>

表 2 提案文の出力例の一部

ベースライン	このスマートフォンは、最新の画像処理エンジンと大型センサーを搭載しているため、写真の美しさや細かい部分までしっかりと表現できます。まるでプロのカメラのような高い解像度で、海外旅行やショッピングの際も思い出やお気に入りの瞬間を鮮やかに残すことができます。
提案手法	このスマートフォンは、一眼レフのような美しい写真が簡単に撮れるだけでなく、大型センサーと最新の画像処理技術によって、細部まで鮮やかに再現できるモデルです。例えば、海外旅行で訪れた絶景の風景や、百貨店で見つけたお気に入りのアイテムを、誰もが羨むような高画質で撮影して、SNSで自信を持ってシェアできます。

表 3 各手法における文章類似度

	平均類似度
ベースライン	0.65 ± 0.1004
提案手法	0.64 ± 0.1028

興味関心があり、流行に敏感な傾向のあるユーザーのパーソナルデータを与えたものである。ベースラインの提案文では、端末のカメラ性能に関する機能的な説明に限定されており、利用方法への言及も“海外旅行やショッピングの際も思い出やお気に入りの瞬間を鮮やかに残す”といった抽象的な表現に留まっている。それに対して提案手法は、“海外旅行で訪れた絶景の風景”や“百貨店で見つけたお気に入りのアイテム”といった具体的な利用シーンの描写に加え、“SNSで自信を持ってシェアできます”のように、体験を通じてユーザーが具体的にどのようなメリットを享受できるかまで解釈している点に顕著な差異がある。このように、ユーザーが自分ごととして認識しやすい体験イメージを提示することで、商品特徴の羅列のみでは得られない情緒性および説明性を反映できたと考えられる。

5 分析

本研究では、生成された提案文が個々のユーザーにパーソナライズ化されているかを検証するため、文章表現の多様性に関する定量評価を行った。日本語事前学習済みモデルである Sentence-BERTⁱⁱ⁾ [16] を用い、各手法において生成された全 200 件の提案文を分散表現に変換し、その相互間の余弦類似度の平均値を算出した。算出結果を表 3 に示す。ベースラインのスコアが 0.65 であったのに対し、提案手法は 0.64 とわずかに低い値を示した。この結果から、同じ端末を提案するため類似度は近くなる傾向がある条件下であっても、提案手法の方が体験イメージ

に基づいた多様な表現を生成できていることが推察される。

6 おわりに

本研究では、推薦商品に対する満足度の高い意思決定および購買体験を実現するため、ユーザー個人のライフスタイルに基づいた体験価値を物語形式で提示する新たなレコメンデーション手法を提案した。提案手法では、ユーザー情報の分析から提案時の戦略策定までを CoT で段階的に実行した。評価実験の結果、LLM による自動評価およびアナテーターによる人手評価の両方において、提案手法は多くの項目でベースラインを上回る結果を示し、利用シーン等の体験イメージを付与した提案が納得感のある意思決定を支援する上で有効な手法であることを確認した。今後の課題としては、より多様な商品ドメインへの適用可能性や、パーソナルデータから得られる潜在的な要望と対話から得られたユーザーの顕在的な要望に差異が発生している場合の検証が挙げられる。

参考文献

- [1] David Goldberg, David A. Nichols, Brian M. Oki and Douglas B. Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 1992. Vol. 35, No. 12, pp. 61-70.
- [2] Yunfan Gao, Tao Sheng, Youlin Xiang, Yun Xiong, Haofen Wang and Jiawei Zhang. Chat-REC: Towards Interactive and Explainable LLMs-Augmented Recommender System. *ArXiv abs/2303.14524*, 2023.
- [3] Ítallo Silva, Leandro Marinho, Alan Said, and Martijn C. Willemsen. Leveraging ChatGPT for Automated Human-centered Explanations in Recommender Systems. In *Proceedings of the 29th*

ⁱⁱ⁾ <https://huggingface.co/pkshatech/GLuCoSE-base-ja>

- International Conference on Intelligent User Interfaces**, 2024. pp. 597–608.
- [4] Sebastian Lubos, Thi Ngoc Trang Tran, Alexander Felfernig, Seda Polat Erdeniz, and Viet-Man Le. LLM-generated Explanations for Recommender Systems. **In Adjunct Proceedings of the 32nd ACM Conference on User Modeling**, 2024. pp. 276–285.
- [5] Yuanjun Feng, Stefan Feuerriegel and Yash Raj Shrestha. Contextualizing Recommendation Explanations with LLMs: A User Study. **ArXiv abs/2501.12152**, 2025.
- [6] Qiyao Peng, Hongtao Liu, Hua Huang, Jian Yang, Qing Yang, and Minglai Shao. A Survey on LLM-powered Agents for Recommender Systems. **In Findings of the Association for Computational Linguistics: EMNLP**, 2025. pp. 11574–11583.
- [7] Yancheng Wang, Ziyang Jiang, Zheng Chen, Fan Yang, Yingxue Zhou, Eunah Cho, Xing Fan, Yanbin Lu, Xiaojiang Huang, and Yingzhen Yang. RecMind: Large Language Model Powered Agent For Recommendation. **In Findings of the Association for Computational Linguistics: NAACL**, 2024. pp. 4351–4364.
- [8] Yubo Shu, Haonan Zhang, Hansu Gu, Peng Zhang, Tun Lu, Dongsheng Li and Ning Gu. RAH! RecSys–Assistant–Human: A Human-Centered Recommendation Framework With LLM Agents. **IEEE Transactions on Computational Social Systems** **11**, 2023. pp. 6759-6770.
- [9] Xu Huang, Jianxun Lian, Yuxuan Lei, Jing Yao, Defu Lian and Xing XieJianxun. Recommender AI Agent: Integrating Large Language Models for Interactive Recommendations. **ACM Transactions on Information Systems** **43**, 2023. pp. 1-33.
- [10] Jizhi Zhang, Keqin Bao, Wenjie Wang, Yang Zhang, Wentao Shi, Wanhong Xu, Fuli Feng, and Tat-Seng Chua. Prospect personalized recommendation on large language model-based agent platform. **arXiv preprint arXiv:2402.18240**, 2024.
- [11] An Zhang, Yuxin Chen, Leheng Sheng, Xiang Wang, and Tat-Seng Chua. On Generative Agents in Recommendation. **In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24)**, 2024. pp. 1807–1817.
- [12] Lei Wang, Jingsen Zhang, Hao Yang, Zhi-Yuan Chen, Jiakai Tang, Zeyu Zhang, Xu Chen, Yankai Lin, Hao Sun, Ruihua Song, Xin Zhao, Jun Xu, Zhicheng Dou, Jun Wang, and Ji-Rong Wen. User Behavior Simulation with Large Language Model-based Agents. **ACM Transactions on Information Systems** **43**, 2025. pp. 1-37.
- [13] Yu Shang, Peijie Liu, Yuwei Yan, Zijing Wu, Leheng Sheng, Yuanqing Yu, Chumeng Jiang, An Zhang, Fengli Xu, Yu Wang, Min Zhang and Yong Li. AgentRecBench: Benchmarking LLM Agent-based Personalized Recommender Systems. **ArXiv abs/2505.19623**, 2025.
- [14] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed H. Chi, F. Xia, Quoc Le and Denny Zhou. Chain of Thought Prompting Elicits Reasoning in Large Language Models. **ArXiv abs/2201.11903**, 2022.
- [15] Krisztian Balog and Filip Radlinski. Measuring Recommendation Explanation Quality: The Conflicting Goals of Explanations. **In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval**, 2020. pp. 329–338.
- [16] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. **ArXiv abs/1908.10084**, 2019.

A 付録

表 4 評価基準

カテゴリ	評価指標	説明
情緒性 価値	有効性	自分にとって「良い選択」であるか判断できそうですか？
	説得力	その商品を試してみよう（受け入れよう）という気になりましたか？
	満足度	提案文を含めた全体的な体験として、好感が持てますか？
	共感性	あなたの現在の状況や悩みに寄り添い、「自分のことを分かってくれている」と感じられましたか？
	発見性	自分では言葉にできなかったニーズや、意外だけど魅力的な選択肢に気づくことができましたか？
説明性 評価	透明性	どのような基準で商品がお勧めされたのかが理解できますか？
	信頼性	おすすめされた内容を信頼できると感じますか？
	精査可能性	好みが正しく理解されているか（あるいは間違っているか）を確認できそうですか？
	効率性	自分がその商品を好きになれるかどうかを「素早く」判断するのに役立ちますか？
	視覚化可能性	その商品に関連した映像や感覚（色、形、音、感触など）が、脳内で鮮明にイメージできますか？