

仮想ユーザ像を用いた情報推薦インタフェースの LLM を用いた評価に関する予備的検討

Preliminary Study on LLM-based Evaluation of Recommender Interface Employing Virtual Users

山崎 洋紀^{1*} 柴田 祐樹¹ 高間 康史¹

Hironori Yamazaki¹, Hiroki Shibata¹, Yasufumi Takama¹

¹ 東京都立大学大学院 システムデザイン研究科

¹Graduate School of Systems Design, Tokyo Metropolitan University

Abstract: This paper investigates the use of Large Language Model (LLM) instead of human participants to evaluate the recommender interface employing virtual users. Recommender systems employing virtual users have been proposed as an explainable recommendation without collecting information from users. As a means for preliminary considering the interface design, this paper reports a result of an experiment, in which LLM decides what movies to watch using the prototype interface.

1 はじめに

本稿では、仮想ユーザ像を用いた情報推薦インタフェースの評価に LLM を用いるアプローチを提案し、予備的検討を行った結果を報告する。

情報推薦システムは、大量に蓄積された情報から必要な情報を効率よく得るために重要な存在となっている。ユーザから嗜好に関する情報を収集し、それらの情報に基づいて推薦アイテムを選択する。しかし、このような個人化された推薦システムに対し、プライバシーが侵害されていると感じるユーザもいるため、システムへの信頼低下につながる可能性が指摘されている[1]。

推薦の受容性を高める方法として、説明可能情報推薦が研究されている。推薦時に、そのアイテムが推薦された理由を説明することで、推薦の透明性や説得力、満足度が向上すると言われている[2]。

これらの背景から、推薦対象ユーザから情報を収集しない説明可能情報推薦を目的として、仮想ユーザ像を用いた情報推薦が提案されている[3, 4]。この手法では、推薦の際に仮想のユーザのプロファイルと、対象アイテムに対する評価値を提示する。ユーザはそれらの情報を参考に推薦アイテムを受容するかどうかを判断するため、ユーザの嗜好に関する情

報を収集する必要がないという利点がある。

仮想ユーザ像を用いた情報推薦インタフェースにおいて、仮想ユーザの人数が少ない場合、提示された仮想ユーザ像が対象アイテムを受容するかどうかの判断の手がかりとして不足である可能性がある。そのため、多様なユーザにとって有用な情報推薦システムとするために、多くの仮想ユーザ像を用いるべきであると考えられる。しかし、多数の仮想ユーザ像を全て同時に提示すると情報過多につながり、仮想ユーザ像を参考にすることが困難になる。従って、同時に提示すべき仮想ユーザ数や、提示すべき情報の内容、ユーザが可能な操作など、ユーザインタフェースの設計において検討すべき事項は多岐に渡る。

インタフェースの評価は通常、実験協力者によるユーザ実験に基づき行うことが一般的であるが、実験協力者を集めるコストや、実験協力者の負担などのため、多数の要素を検討することは困難である。

そこで本稿では、仮想ユーザ像を用いた情報推薦インタフェースの初期段階での検討に、大規模言語モデル (Large Language Model: LLM) を実験協力者の代わりに用いることを検討する。LLM を用いて推薦システムを利用するユーザの行動をシミュレートする手法は、実際のユーザ実験の低コストな代替手段になる可能性として注目されている[5, 6]。既存研究ではユーザのアイテム評価行動を LLM に代替させることがほとんどであるのに対し、本稿ではインタフェースの操作を LLM に行わせ、その結果をインタフェースの評価・改善に利用する点で異なる。

*連絡先: 東京都立大学大学院 システムデザイン研究科

〒191-0065 東京都日野市旭が丘 6-6

E-mail: yamazaki-hironori@ed.tmu.ac.jp



図 1: 仮想ユーザを評価値別に提示する
 インタフェース



図 2: プロファイル文を表示した画面

LLM によるインタフェース評価の予備的検討として、インタフェースで提示する情報、および可能な操作をプロンプトで LLM に与え、提示された映画を視聴するか否かを判断するタスクをさせる実験を行った。LLM が判断の手がかりとした情報を分析し、インタフェース改善に有益な情報が得られるか、および評価方法の改善点について考察する。

2 関連研究

2.1 仮想ユーザ像を用いた情報推薦

稲田らは、推薦対象ユーザの情報を利用しない説明可能情報推薦を目的として、仮想ユーザ像を用いた情報推薦手法を提案している[3, 4]。ユーザ像はポジティブおよびネガティブな嗜好の説明を 3 文ずつの計 6 文から構成されるプロフィール文として生成し、人手による生成手法と LLM を用いた半自動生成手法を提案している。生成には匿名ユーザの行動履歴データを用いている。



図 3: 仮想ユーザの情報を確認できる画面

2.2 LLM を用いた推薦システムの評価

推薦システムを利用するユーザの行動を、LLM を用いてシミュレーションする研究が行われている。Agent4Rec[5]や RecAgent[6]は、ユーザのプロファイル、過去の行動、推薦システム内で行うことができる行動をプロンプトに入力し、ユーザと推薦システムのやり取りをシミュレートしている。実際の人間の行動をどの程度再現できるかについて検討し、嗜好を高い精度で反映させることや、フィルターバブル現象を再現できることを確認している。

3 提案手法

3.1 仮想ユーザ像を用いた情報推薦インタフェース

本節では、評価対象として想定する、仮想ユーザ像を用いたインタフェースについて説明する。このインタフェースでは、映画を対象アイテムとしており、提示された映画に対する評価値別に数名の仮想ユーザを提示する。対象アイテムに対し、高評価あるいは低評価をつける仮想ユーザの嗜好の傾向を把握し、自身の嗜好と比較することで、提示されたアイテムを選択するか否かを判断することを想定して

現在のページ

映画タイトル : Toy Story

概要ページのリンク :

IMDB

<http://www.imdb.com/title/tt0114709/>

allcinema

<https://www.allcinema.net/cinema/28621>

旬報キネマ

なし

仮想ユーザの情報

以下は提示された仮想ユーザの情報です。

仮想ユーザは評価値別に提示されています。

プロフィール文が未確認の仮想ユーザは、[未確認]と表記されます。

評価 5 をつけた仮想ユーザ

ユーザ名 : D

プロフィール : 個性的で挑発的な作品を評価 : 「Pulp Fiction (1994)」や「Fight Club (1999)」のように、斬新なストーリーテリングや強烈なテーマに惹かれる。

(中略)

このユーザが映画につけた平均評価値 : 3.46

高評価した映画 (この仮想ユーザが特に好むと判断した映画) :

Night Tide

(中略)

低評価した映画 (この仮想ユーザが好まないと判断した映画) :

Porky's II: The Next Day

(中略)

ユーザ名 : U

プロフィール : [未確認]

評価 4 をつけた仮想ユーザ

ユーザ名 : J

プロフィール : [未確認]

(中略)

評価 1 をつけた仮想ユーザ

なし

いる。

図 1 に、インタフェースのスクリーンショットを示す。映画のタイトル、概要ページのリンクと、1~5 の 5 段階の各評価値について最大 3 名の仮想ユーザの名前が表示される。概要ページとして、allcinema¹、キネマ旬報 WEB²、IMDb³を利用する。仮想ユーザの名前をクリックすると、その嗜好を表すプロフィール文が表示される (図 2)。この時、仮想ユーザの名前を変更するボタンと仮想ユーザの情報を確認できる画面へのリンクが同時に表示される。仮想ユーザの名前を変更するボタンをクリックすると、仮想ユーザの名前を自由に変更することができる。この機能は、仮想ユーザの識別性を向上させることを目的としている。

仮想ユーザの情報を確認できる画面を図 3 に示す。この画面では、仮想ユーザの現在の名前、元の名前、平均評価値、プロフィール文、その仮想ユーザが高評価した映画と低評価した映画が表示される。仮想ユーザが高評価した映画と低評価した映画はランダムにそれぞれ 5 本選ばれる。

3.2 LLM を利用したインタフェース評価実験

LLM にインタフェースを操作させるため、インタフェースで提示する情報などをテキストとして LLM に入力し、実行する操作を出力させる。評価実験の手順は以下の通りである。

1. LLM にプロンプトを入力する
2. LLM が行動を選択して出力する
3. 選択された行動に従いプロンプトを更新し、1 に戻る

LLM に入力するプロンプトには、システムに関する説明、LLM が模擬するユーザの嗜好に関する情報、インタフェースが提示している情報、これまでに行った行動、思考プロセス、次に行うことができる行動を記述する。

インタフェースが提示している情報の例を図 4 に示す。図 1 に示す画面をベースとしており、映画のタイトル、概要ページのリンク、仮想ユーザの情報から構成される。仮想ユーザの情報は、評価値毎に仮想ユーザの名前とプロフィール文を記述するが、最初はプロフィール文は「未確認」とし、内容はブ

¹ <https://www.allcinema.net/>

² <https://kinejun.jp/>

³ <https://www.imdb.com/>

図 4: インタフェースが提示している情報の記述例

思考プロセス

あなたが映画を視聴するかどうかを決定する前に、以下の思考プロセスを順に実行してください。

1. あなたの映画の嗜好と、提示された映画との適合性はどうですか？
2. 仮想ユーザのプロファイルと、彼らがこの映画にその評価をつけた理由は、あなたの意思決定にどのような影響を与えますか？
3. 未確認の仮想ユーザのプロファイルが視聴決定に影響を与えると考えるなら、[CHECK_PROFILE]コマンドでその情報を確認することができます。
4. 確認した概要ページがある場合、該当する概要ページのリンク先の情報を意思決定に反映させることができます。あなたが確認した概要ページは、行動履歴に記載されています。まだ確認していない概要ページから映画の概要を確認したい場合は、[CHECK_MOVIE_INFO]コマンドで概要ページを確認することができます。
5. 総合的に判断し、最終的な視聴決定とその理由を導き出してください。
6. もし今後、あなたが特定の仮想ユーザを分かりやすく識別したいと考えるなら、仮想ユーザの名前を変更することができます。

図 5: 思考プロセスの記述例

プロンプトに含めない。「未確認の仮想ユーザのプロファイル文を確認する」という操作を実行した後に、その仮想ユーザのプロファイル文を含めて記述する。

思考プロセスの記述例を図 5 に示す。映画を視聴するか否かを決定する前に実行する思考プロセスを記述する。自身の嗜好と映画との適合性、仮想ユーザの情報を考慮し、必要であれば仮想ユーザのプロファイル文や概要ページを確認するように指示する。

次に行うことができる行動として、以下の 6 つの選択肢を提示し、LLM にどの行動を選択するかを出力させる。

- 映画の視聴を決定する。
- 映画を視聴しないことを決定する。
- 未確認の仮想ユーザのプロファイル文を確認する。
- 映画の概要ページを確認する。
- 特定の仮想ユーザが高評価・低評価した映画や平均評価値を確認する。
- 仮想ユーザの名前を変更する。

表 1: LLM が模擬するユーザの嗜好情報の例

項目	内容
人物像	スリラー・サスペンス愛好家
映画の嗜好	予想外の展開や緊張感のあるスリラー映画、心理的な駆け引きを描いた作品が好き。
好む映画	『セブン』 『シャッター アイランド』 『プリズナーズ』
好まない映画	恋愛映画やコメディ映画、ストーリー展開が単純な作品

未確認の仮想ユーザのプロファイル文を確認する行動は、仮想ユーザの名前をクリックしプロファイル文を表示する操作に対応する。前述の様に、この行動が選択された場合は、以降のプロンプトからプロファイル文は「未確認」ではなく内容を記述する。

映画の概要ページを確認する行動は、概要ページのリンクをクリックする操作に対応する。この行動が選択されると、行動履歴に概要ページを閲覧したことが追記される。図 5 に記載がある通り、行動履歴に概要ページを確認したことが記述されていれば、LLM はリンク先の情報を意思決定に反映させることができる。

特定の仮想ユーザが高評価・低評価した映画や平均評価値を確認する行動は、仮想ユーザの情報を確認できる画面への遷移に対応する。この画面で確認できる、仮想ユーザが映画につけた平均評価値、高評価、低評価した映画それぞれ 5 本が、図 4 のユーザ D のようにプロンプトに含まれるようになる。

仮想ユーザの名前を変更する行動は、該当機能を利用する操作に対応する。この行動を選択する際は、新しい名前を同時に出力させ、以降はその仮想ユーザは新しい名前でもプロンプトに記述される。

映画の視聴可否の決定が選択された場合、インタフェース操作を終了し、インタフェース利用の感想、仮想ユーザの参考度、対象映画を視聴後の評価を出力するよう指示するプロンプトを入力する。

4 評価実験

4.1 実験概要

3 節で述べた手順に従い、LLM に提案インタフェースのプロトタイプを評価させる実験を行った。映

表 2: 視聴可否判断の根拠

模擬 ユー ザ	視聴する				視聴しない				計
	T1	T2	A1	A2	T1	T2	A1	A2	
概要	1	1	2	2	1	0	2	1	10
評価 分布	0	1	1	0	1	1	0	0	4
高評 価	1	1	3	3	4	2	2	0	16
低評 価	0	1	0	0	1	0	0	1	3
判断 材料 不足					0	3	1	0	4
作品 数	1	2	3	3	4	3	2	2	20

画に関する情報および仮想ユーザ像の生成には MovieLens の 1m データセット⁴を用いた。仮想ユーザ像は稲田ら[3]の LLM を用いた半自動生成手法によって 22 名生成した。

LLM が模擬するユーザの嗜好情報の例を表 1 に示す。ChatGPT⁵を用いて生成し、人物像、映画の嗜好、好む映画、好まない映画から構成される。「スリラー・サスペンス愛好家」を 2 名 (T1, T2) と、「アニメ・ファンタジー好き」を 2 名 (A1, A2) 用意した。同じジャンルを好むユーザでも、嗜好の詳細や好む映画などが異なるように生成している。

模擬するユーザ 1 名につき 5 本の映画について実験を行った。映画はランダムに 5 本選んだが、好むジャンルが同じ 2 名については同じ 5 本の映画について実験した。

インタフェース操作を行う LLM は OpenAI⁶の GP T-5.2 モデルを使用した。

4.2 結果

模擬ユーザが選択可能な 6 種類の行動のうち、映画の概要ページの確認と仮想ユーザのプロファイル文の確認は、20 回の試行全てにおいて行われた。一方、仮想ユーザが高評価・低評価した映画や平均評価値の確認は 1 回しか選択されず、仮想ユーザの名前の変更は 1 回も行われなかった。

⁴ <https://grouplens.org/datasets/movielens/>

⁵ <https://chatgpt.com/> (2025 年 1 月 17 日, 2026 年 1 月 16 日)

⁶ <https://openai.com/>

4.2.1 視聴可否の判断理由

LLM が視聴するかどうかを判断した理由を分析したところ、判断の根拠は映画の概要、仮想ユーザの評価分布、高評価・低評価それぞれの仮想ユーザと自身の嗜好の比較に大別された。また、判断材料が不足しているため視聴しないと判断したとの回答もあった。これらの根拠が主な判断理由として挙げられた回数を、視聴する・しないと判断した場合毎に集計した結果を表 2 に示す。複数の理由を挙げた場合はそれぞれカウントしている。前述のとおり、仮想ユーザが評価した映画を確認する行動は観測されなかったが、プロファイル文を根拠とした場合でも、仮想ユーザが好む映画などの具体例に言及することはなかった。

表 2 より、映画の概要と、高評価をつけている仮想ユーザを根拠とするケースが多い結果となった。一方で、低評価した仮想ユーザが根拠となることは少なかった。また、映画によっては、提示された仮想ユーザの評価分布が、高評価あるいは低評価のどちらかのみの場合があり、視聴判断に影響を与えていた。さらに、判断材料不足は模擬ユーザ T2 によくみられた。

この結果から、高評価している仮想ユーザのプロファイル文は初めから表示し、仮想ユーザの特徴把握の手間を軽減するといった改善案が考えられる。仮想ユーザの評価が高評価・低評価のどちらかに偏る問題については、仮想ユーザ数を増やせば起こりにくくなると想定される。しかし、非常に多数の仮想ユーザを用いているにも関わらず評価に偏りがある場合には、どのような嗜好のユーザにとっても似た評価になる映画と考えられるため、「万人受けする映画」などの説明を記載することで視聴判断の助けになる可能性がある。また、自身の嗜好と比較できる材料が不足しているという理由が挙げられていたことから、インタフェースが提示する情報もしくは仮想ユーザ像が不足している可能性もある。自分の嗜好に関する手がかりを得られないという状況を起こさないためには、それぞれ異なる嗜好を持つ、多様な仮想ユーザ像を提示することが効果的と考える。

4.2.2 低評価仮想ユーザの有無

実験に用いた 10 本の映画のうち 5 本について、提示された全ての仮想ユーザのプロファイル文が確認された。提示された仮想ユーザの一部のみプロファイル文を確認して視聴可否を判断したケースでは、低評価している仮想ユーザがいなかった。全ての仮想ユーザのプロファイル文が確認された 5 本の映画のうち、低評価している仮想ユーザがいなかった映画は 1 本であり、その場合は評価 3, 4 の仮想ユーザしか存在しなかった。

表 3. 仮想ユーザのプロファイル文の極性と解釈

		高評価		低評価	
		P	N	P	N
好む要素	あり	6	0	0	0
	なし	8	0	1	3
好まない要素	あり	0	0	0	0
	なし	1	3	0	0
計		15	3	1	3

この結果から、評価が割れている場合は両方の意見を慎重に確認しているといえるが、4.2.1 節で述べた通り、低評価をつけている仮想ユーザは判断の根拠となることが少ない。この理由について考察するため、模擬ユーザが判断理由において、自身の好む・好まない要素が含まれるかどうかについて、仮想ユーザのプロファイル文を解釈した回数を集計した結果を表 3 に示す。表において、行は好む・好まない要素がプロファイル文に含まれるか否かに対応し、列は対象の映画を高評価・低評価した仮想ユーザに対応する。また、プロファイル文は仮想ユーザのポジティブ (P)、ネガティブ (N) な嗜好の両方を記載しているため、それぞれ分けて集計している。

表 3 より、高評価をつけている仮想ユーザのポジティブな嗜好から、好む要素の有無を解釈する回数が最も多く、結果として視聴判断の手がかりとされることが多かったと考える。一方で、低評価をつけている仮想ユーザはネガティブな嗜好の方が参考にされており、好む要素を含まないという解釈に使われている。好まない要素が含まれているかどうかは概要ページで確認していたため、低評価ユーザを手がかりとすることが少なかったと考える。

模擬ユーザが言及したプロファイル文をみると、映画のジャンルや雰囲気を示す言葉を含む文(「スタイリッシュで知的な作品を評価」「奇抜なホラーや独創的な世界観に魅了される」等)や、特定の要素に対するネガティブなプロファイル文(「感傷的すぎる映画には厳しい」等)が言及されやすかった。一方で、「時代を超えたクラシック作品を評価」や「平凡な娯楽映画に興味薄い」など具体性に欠ける賛辞・批判のプロファイル文や、「期待を裏切るシリーズ作に失望」など限定的すぎるプロファイル文は言及されなかった。

この結果から、映画のジャンルや雰囲気に直結しやすいポジティブな嗜好を持つ仮想ユーザや、好まない要素が明確な仮想ユーザが手がかりになりやすいといえる。従って、図 1 で仮想ユーザを評価値別に提示する際に、このような特徴を満たすプロファイル文を持つ仮想ユーザを強調表示するといった方

法が考えられる。

5 おわりに

本稿では、仮想ユーザ像を用いた情報推薦インタフェースの評価に LLM を用いるアプローチを提案した。予備的検討として、提示した映画の視聴可否を LLM に判断させる実験を行った結果、高評価をつけている仮想ユーザの重要性や手がかりとなりやすい仮想ユーザの特徴などの知見が得られた。今後は、各仮想ユーザの特性・役割の違いについて詳細に分析する他、本実験で得られた知見に基づいて、提案するインタフェースの評価実験を行う予定である。

謝辞

本研究の一部は JSPS 科研費 22K19836, 23K24953 の助成を受けたものです。

参考文献

- [1] F. Ricci, L. Rokach, B. Shapira: Introduction to Recommender Systems Handbook, Springer, pp. 1-35, 2010
- [2] Y. Zhang, X. Chen: Explainable Recommendation: A Survey and New Perspectives, Foundations and Trends in Information Retrieval, Vol. 14, No. 1, pp. 1-101, 2020
- [3] 稲田真樹人, 柴田祐樹, 高間康史: 仮想ユーザ像を用いた説明可能情報推薦にむけた予備的検討, 第 39 回ファジィシステムシンポジウム講演論文集, No. 39, pp. 553-558, 2023
- [4] 高間康史, 柴田祐樹: 仮想ユーザ像を用いた情報推薦システムにおける大規模言語モデルを用いたプロファイル生成の検討, 第 39 回人工知能学生全国大会, No. 1E4-OS-3a-04, 2025
- [5] A. Zhang, Y. Chen, L. Sheng, X. Wang, T. S. Chua: On Generative Agents in Recommendation, 47th International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 1807-1817, 2024
- [6] L. Wang, J. Zhang, H. Yang, Z. Y. Chen, J. Tang, et al.: User Behavior Simulation with Large Language Model-based Agents, ACM Transactions on Information Systems, Vol. 43, No. 2, Article 55, pp. 1-37, 2025