

小規模言語モデルが抽出する経験価値に基づく ゲーム推薦システムの構築

長尾 羽留[†] 亀谷 由隆^{††}

[†] 名城大学大学院理工学研究科情報工学専攻 〒468-0073 名古屋市天白区塩釜口一丁目 501 番地

^{††} 名城大学情報工学部情報工学科 〒468-0073 名古屋市天白区塩釜口一丁目 501 番地

E-mail: ^{††}tykameya@meijo-u.ac.jp

あらまし 近年巨大化の一途を辿っているゲーム市場では情報検索の仕組みが必須であるが、従来の検索方法ではゲームの本質を捉えることは困難である。そこで本研究では、ゲームは他のマルチメディアと比較して自らが働きかける側面、すなわちインタラクティブ性が強いということに注目し、プレイヤー自身の経験によって得られる価値（経験価値）が重要であると考えた。具体的には、Steam Web API から収集したユーザーレビューに対し、小規模言語モデルによって同じ主張をしているレビュー同士をグループ化し、その中で共通する観点を経験価値として抽出する。そして、それらの観点に基づく絞り込み条件をユーザが指定できるゲーム推薦システムを提案する。被験者実験では、好意的な評価も多くあったが観点の分かりやすさや全体的なシステムの使いやすさに関して課題が残った。

キーワード ゲーム推薦, 経験価値, レビュー分析, 小規模言語モデル

1 はじめに

近年巨大化の一途を辿っているゲーム市場において、従来のキーワード検索ではゲームの本質を捉えられず、ユーザが求めているゲームを見つけ出すことは困難である。特に Web 検索では、概略の情報や一般的な知識などといった期待しない不要なページを多く含んでしまう。この問題に対処するために本研究では、ゲームは他のマルチメディアと比較して自らが働きかける側面、すなわちインタラクティブ性が強い [12] ということに注目し、ゲームの本質を捉えるには、その商品を購入した人自身の経験によって得られる価値（経験価値）が重要であると考えた。

経験価値とは、1990 年代末にコロンビア・ビジネス・スクールの B. H. Schmitt によって提唱されたマーケティング手法の概念 [9] であり、一般的に 5 つの戦略的経験モジュール (strategic experiential module, SEM) と呼ばれるカテゴリに大きく分類される。その 5 つとは、SENSE (感覚的知覚), FEEL (情緒的感情), THINK (創造的思考), ACT (身体性), RELATE (社会性) である。SENSE は五感、ゲームであれば特に視覚・聴覚の直接的な刺激によって得られる価値である。FEEL は感情の動きによって得られる価値である。これは理性によって制御できない領域である衝動的な感情を扱う。THINK は知的好奇心や創造心の刺激によって得られる価値である。ACT は他者との接触が生じたときや、生理的な欲求、さらにはライフスタイルの変化によって得られる。RELATE では他者との繋がりを求める欲求を満たすことや、単に所属することによる安心によって得られる。

中谷ら [7] は、事前に人手で作成した辞書を用いて経験価値をレビュー文から抽出し、ゲームの経験を定量化し経験値へと変換させ、個々のユーザの経験に適するゲームを推薦するシス

テムを提案した。その中で、経験価値の取得にユーザーレビューを選択した。その理由は、ユーザーレビューには“画像がとても汚い”などの主観的な評価や“何度も泣いてしまった”などの感情的な文章が含まれていて、これらはゲームの経験価値を表している、ということである。

本研究では、Steam Web API から収集したユーザーレビューを対象に小規模言語モデル (small language model, SLM) Qwen3-8B¹ を用いて、ゲーム体験に関する経験価値の観点を抽出する。提案手法では各レビューから観点を抽出し、同一または類似の主張を持つ観点同士を段階的にグループ化する。このレビュー集約を繰り返すことで詳細な観点グループを徐々に一般化させ、最終的にユーザが選択可能な観点名を出力する。この得られた観点を絞り込み条件を指定できるゲーム推薦システムを提案する。

経験価値を観点として抽出することは、豊田ら [10] の研究から着想を得ている。豊田らは、食べログのレビューから Google Cloud Platform の Natural Language API² を用いてエンティティを抽出し、エンティティの分散表現に基づいてクラスタリングしたエンティティ集合から、コサイン類似度に基づき各クラスタの代表語を観点として抽出した。そして、それらの観点を絞り込み条件としてインタフェース表示することで、飲食店の特徴を捉えられる推薦システムを提案した。

また、レビュー集約は中井ら [6] の研究からも着想を得ている。中井らは、大規模言語モデル (large language model, LLM) を用いてレビューから観点と評価を抽出し、類似している観点・評価のクラスタリングを行った。そして観点ごとに評価を対応付けた表を作成することで、人の意見を基に短時間で商品比較可能なシステムを提案した。

1 : <https://huggingface.co/Qwen/Qwen3-8B>

2 : <https://cloud.google.com/natural-language?hl=ja>

本論文の構成を示す。まず第2節で先行研究について述べる。次に第3節で提案システムの構成や仕様について述べる。第4節で実施した被験者実験の方法を述べ、示した結果について考察する。最後に第5節で本論文をまとめる。

2 先行研究

ユーザーレビューから抽出したアイテムの特徴量を中心とした推薦システムの開発に関しては、以前から様々な研究がなされている。大山ら [8] は、ゲームレビューサイトに掲載されているレビュー文を word2vec³ に学習させ、言葉の足し引きによってゲームを定量化したゲーム推薦システムの研究を行った。しかし、この大山らの研究、前述の豊田らの研究いずれも経験価値を上手く抽出できていない、単語単位でレビュー分析を行っているために感情極性を考慮できていないなどの課題が挙げられている。

また、近年の深層学習の進展は文章分析や情報抽出 (information extraction, IE) の分野に大きな影響を与えており、非構造化テキストから構造化情報を抽出する上で重要な役割を果たしている。例えば Wu ら [11] は、エンティティの予測集合と正解集合の類似度を測る柔軟な評価指標 AESOP (Approximate Entity Set Overlap) の提案および、出力トークン数削減と多段階処理による効率性と精度の向上を実現した言語モデル MuSEE (Multi-Stage Structured Entity Extraction) の開発を通して、従来の IE よりも多角的な視点から洞察を得られるような枠組み SEE (Structured Entity Extraction) を提案した。Korikov ら [5] は、ユーザクエリを LLM によって複数の観点に自動的に分解し、レビュー文書を観点ごとに順位付けしたうえで統合する Aspect Fusion を提案した。これにより、従来の単一クエリ処理では捉えにくい多観点クエリをもつ検索タスクにおいて高い再現性と精度を実現した。

3 提案システム

本節では、提案システムについて説明する。まず 3.1 節で提案システム全体の構成について述べる。次に 3.2 節で用いたデータセットについて述べる。さらに 3.3 節でレビュー集約プロセスと各プロセスの結果について述べる。

3.1 システムの構成

後述する 3.3 節で作成した観点データを利用して Web 推薦システムを実装した。実装には Python の Web 開発用フレームワークである Flask [4] を用いた。各ゲームに関する概要や動画などの情報は、レビューと同様に Steam Web API のゲームの詳細情報を取得するリクエスト [13] を用いて取得した。図 1 は実装したシステムの構成図である。

図 1 の通り、Steam Web API からのレビュー取得およびレビュー集約はシステムを利用するための前処理段階であり、レビュー分析結果を Flask アプリケーションに保存してからユー

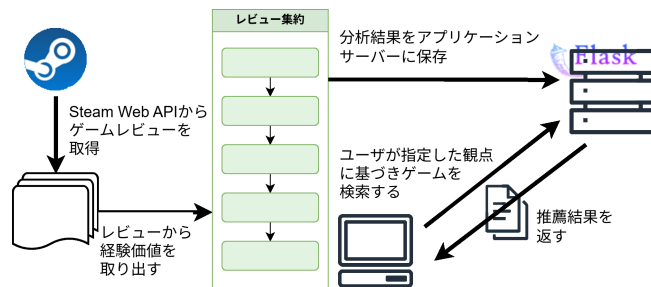


図 1 提案システムの構成図

ザが使用できる状態になる。図 2, 3 はそれぞれユーザー側から確認できる検索画面と検索結果画面である。

検索画面は図 2 の通り、左側の観点エリアと右上の選択観点表示エリア、右下の説明文エリアに分かれる。観点エリアでは抽出された観点が 1 節で述べた SEM およびサブグループごとに表示されている。サブグループは折り畳みのユーザーインターフェース (以下、単に UI) になっており、類似評価グループは絞り込み条件としてチェックボックスで選択できる。ここで、サブグループおよび類似評価グループは中井らの先行研究 [6] に基づいて設定したグループ階層であり、SEM → サブグループ → 類似評価グループの階層関係で構成されている。この詳細については 3.3 節で述べる。ユーザーの選択した観点が右上のエリアに表示されている状態で「観点で検索」ボタンを押すことでゲーム推薦を受けられる。また、サブグループ・観点にマウスを合わせると右下のエリアに説明文が表示される。

検索画面は豊田ら [10] の研究に基づいてレイアウトを設計した。豊田らは実装した飲食店推薦システムの中で、レビューから抽出した観点を「利用シーン」「味」「料理名」「お店の特徴・雰囲気」「その他」の 5 つのカテゴリに分類し、各カテゴリ内の観点をチェックボックスで選択できるような UI を実現している。

検索結果画面は図 3 の通り、左側の推薦ゲームリストエリアと右側のゲーム詳細情報エリアに分かれている。ゲームリストエリアでは、各ゲームのタイトルとヘッダー画像、選択した観点のうち一致するものの一部を 1 つのパネルに表示する。このパネルのいずれかをクリックすると詳細情報エリアが書き換わる。詳細情報エリアでは、ゲームの説明や動画・スクリーンショット、ユーザーが選択した観点のうち一致したものを表示する。

3.2 データセット

本研究では、Steam が公式で提供している Steam Web API のレビュー情報取得 URL⁴ にアクセスして得られたユーザーレビューを用いた。しかし、現在 Steam で扱っているゲームを全て用いるのは規模が大きすぎ、また馴染みのないゲームタイトルは評価が難しいため、今回は AAA タイトルを輩出している企業や日本の有名企業から発売しているゲームタイトルの中で、レビュー数が 10 以上あるアイテムのみに絞り込みを行った。そ

3 : <https://radimrehurek.com/gensim/models/word2vec.html>

4 : <https://partner.steamgames.com/doc/store/getreviews>

ゲーム検索

The screenshot shows a search interface with two main filter columns: SENSE (五感, 特に視覚・聴覚に関する価値) and FEEL (感情に関する価値). The SENSE column includes options like 'Dark Visual Story' and 'Action and Immersion Tutorial'. The FEEL column includes 'Expression and Experience' and 'Core of Gameplay'. To the right, a 'Selected Viewpoint' section shows 'None Selected'. Below that, a 'Viewpoint Search' button is present. The 'Viewpoint Description' section for 'Dark Visual Story' provides a detailed explanation of the game's themes and visual elements.

図2 検索画面のレイアウト

ゲーム一覧

The screenshot displays a game list on the left with cards for 'Puyo Puyo Tetris' and 'Stellar Blade'. The main area shows a detailed view for the game 'Different Stars' (違う星のぼくら), featuring a video player, a list of shared viewpoints (e.g., 'High Difficulty Puzzle', 'Unique Puzzle Mechanics'), and a 'Game Information' section. A 'Game List Area' label is overlaid on the bottom left of the game list.

図3 検索結果画面のレイアウト (画像の出展: Steam)

の結果、タイトル数は1208、全レビュー数は9196となった。

3.3 レビュー集約処理

本節では、収集したレビューデータから観点を抽出し提案システムに適した形に加工する処理(図1の左側)について説明する。これは提案システムが使用可能になるまでの前処理である。ここでは中井らの先行研究[6]を参考に、図4に表される5プロセスからなる処理を行った。5つのプロセスはいずれでもSLM(Qwen3-8B)を利用している。

3.3.1 観点と評価を抽出とSEMへの分類

3.2節で収集したレビューを1つずつQwen3-8Bに入力し経験値を観点と評価という形で抽出した。評価はゲームの経験値を表す部分をレビューから抜き出した1文であり、観点は評価

を端的に表す名前である。図5は、以上の処理の一例である。

3.3.2 観点のサブグループへの集約

プロセス3.3.1で抽出した観点をQwen3-8Bに入力し、似ている主張をしているもの同士を1つのサブグループに集約する。処理のフローは図6の通りであり、階層クラスタリングに似た処理を行う。

具体的には、観点リストを固定長(今回は25個)の小集合に区切り、各小集合からQwen3-8Bを用いたクラスタリングによって生成したグループ名を取り出し、1つのグループ名リストを作る。このリストをまた固定長に区切り、前工程と同様にクラスタリングするという処理を繰り返し、徐々に一般的なグループ名にしていく。クラスタリング結果に変化が見られなくなった段階で得られたグループ名を最終的なサブグループと定

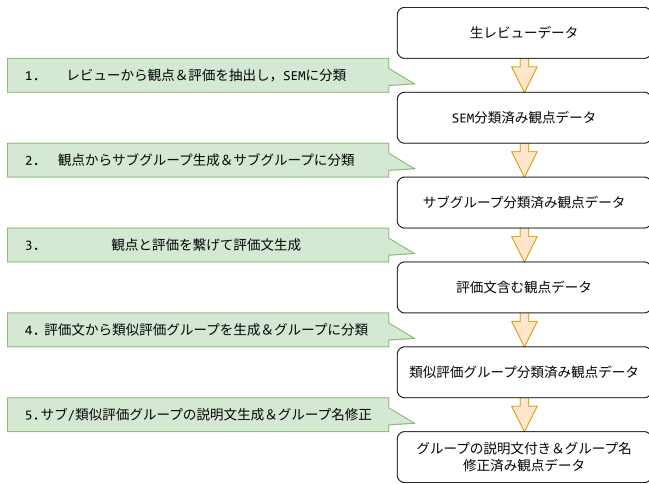


図4 レビュー集約処理の全体図

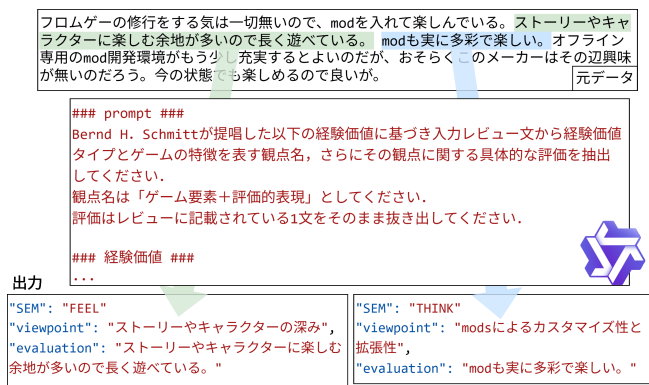


図5 レビューから観点&評価抽出の例

義する。その後、観点を得られたサブグループに分類していく。図7は以上の処理の具体例である。

このような方法を取ったのは、Qwen3-8Bのコンテキスト長上限の中で効果的にレビュー集約するためである。9000件以上のレビューを扱っているとプロンプトがコンテキスト長上限を超えてしまうことが避けられない。Mamba [2] などコンテキスト長上限が高いモデルもあるが、プロンプトが長すぎても抽出精度向上は見込めないと考えた。また、Qwen3の思考モードを効果的に活用するには、クラスタリングは段階的に行うべきであると考えた。この点は、比較的少数のレビューを扱っていたと思われる中井ら [6] の研究では見られなかった視点であると考えている。

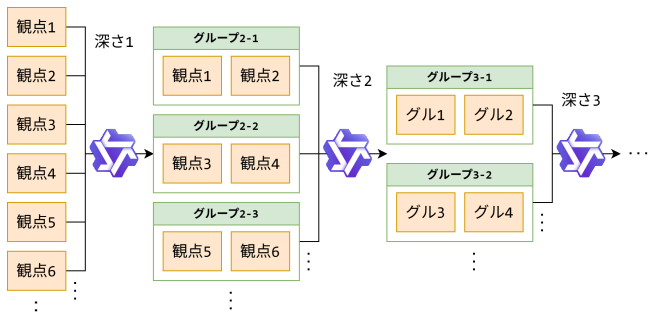


図6 観点のサブグループへの集約フロー

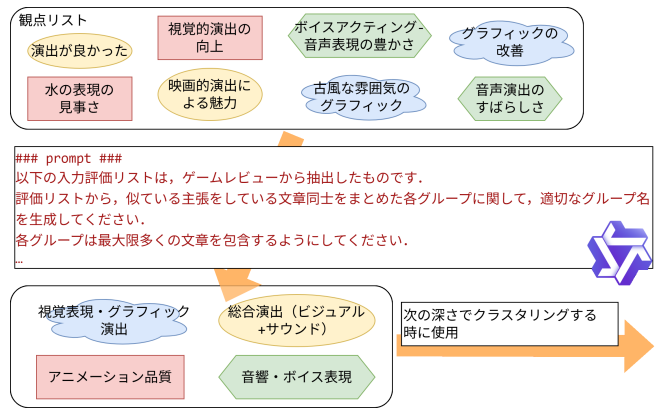


図7 グループ名生成の具体例

3.3.3 評価文の生成

3.3.1節で抽出した観点と評価を繋げ、自然な日本語文を生成する。3.3.2節で得られたサブグループごとに類似評価の集約を行う。評価集約では、各観点グループの中にどのような意見があるのかを簡潔に示すために似ている評価を1文に要約する。しかし、評価のみを使用して要約すると情報が不足する場合がある。例えば、「グラフィック」グループのある観点に対して「いいもの」という評価が抽出された場合、「映像美」に対する「いいもの」という評価と「アートスタイル」に対する「いいもの」という評価は1つにまとめるべきではない [6]。そこで評価を集約するために、観点と評価を1文にまとめた擬似的なレビュー文を作成した。図8は、以上の処理の一例を表している。

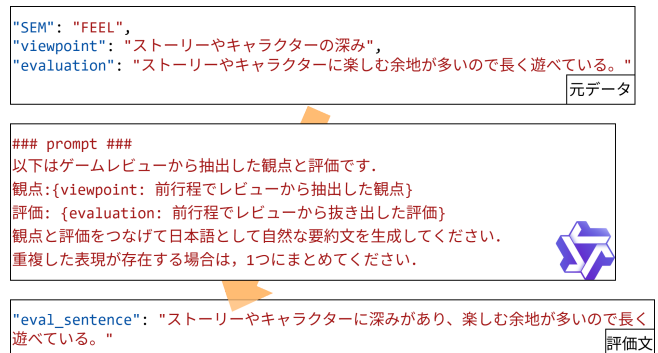


図8 評価文生成の具体例

3.3.4 評価文の類似評価グループへの集約

3.3.3節で生成した評価文を利用して、3.3.2節と同様に同じ主張をしている評価文を集約し、類似評価グループ名を取得する。その後、類似評価グループに分類する。

3.3.5 グループの説明文の生成とグループ名の修正

サブグループ、類似評価グループについて説明文を付与することで、どのような観点がゲームに存在するのかユーザーが把握するのを補助する。その後、得られた説明文に基づいてグループ名をよりユーザーに分かりやすいものに修正する。

4 被験者実験

本節では提案システムを用いた被験者実験の内容と結果，それに対する考察を述べる．情報系の学生 20 名（修士 2 年: 3 人，学部 4 年: 8 人，学部 3 年: 9 人）の協力を得て提案システムを用いた実験を行い，結果を記録した．

被験者実験内容は以下の 2 つである．

ゲーム推薦の評価実験（以下，評価実験 A）

提案システムを用いて，3 つのシナリオに沿ったゲームを探索する．各シナリオ探索終了後に提案システムによるゲーム推薦を評価する質問に回答する．

UI・ユーザビリティの評価実験（以下，評価実験 B）

提案システムのインタフェースおよびユーザビリティに関する質問に回答する．

4.1 ゲーム推薦の評価実験

4.1.1 実験概要

評価実験 A において用いるシナリオは表 1 の 3 つである．被験者は各シナリオに沿って提案システムを用いて遊びたいと感じるゲームを探索する．被験者は各シナリオに沿ったゲーム探索がすべて終了した後に表 2, 3 の 5 段階リッカート尺度における評価の質問に回答する．この 2 つの表は提案システムにて用いた観点を使った推薦において，それぞれ推薦条件と推薦結果に関する評価を尋ねるものである．

表 1～3 の項目は豊田ら [10] の研究に基づいて設定した．

表 1 3 つのゲーム探索シナリオ

No.	シナリオの内容
1	あなたは，美しいビジュアルとサウンドの融合を楽しみつつ，物語に没入できるゲームをプレイしたいと考えています．その条件に当てはまるようなゲームを 1 つ探してください．
2	あなたは，ビジュアルとサウンド，物語の没入性のレベルが高く，多プレイヤーと協力・交流ができるゲームで，マルチプレイでがっつりやり込みたいと考えています．その条件に当てはまるようなゲームを 1 つ探してください．
3	あなたは，ストーリーとキャラクターが丁寧に描かれつつ，挑戦を通じた発見や交流があり，かつコストパフォーマンスが高いゲームをプレイしたいと考えています．その条件に当てはまるようなゲームを 1 つ探してください．

表 2 評価実験 A 推薦条件に関する質問項目

No.	質問内容
1	観点は役に立ったか
2	観点は分かりやすかったか
3	特徴的な観点を発見できたか
4	選択した観点と推薦されたゲームは関連があったか
5	選択した観点はそのゲームの特徴を反映できていると感じたか

各シナリオの表 2, 3 の回答は，ノンパラメトリック検定の

表 3 評価実験 A 推薦結果に関する質問項目

No.	質問内容
1	推薦されたゲームを実際に遊びたいと感じたか
2	推薦されたゲームは気に入ったか
3	推薦されたゲームは見つけにくい良作と感じたか
4	詳細情報画面からそのゲームの特徴が分かったか
5	選択した観点とレビュー文内の該当箇所は関連していたか

一つである一標本の Wilcoxon 符号付き順位検定 (one-sample Wilcoxon signed rank test) を用いた（以下，単に Wilcoxon 符号順位検定と呼ぶ）．

帰無仮説 (H_0)，対立仮説 (H_1) を次のように立てた．

帰無仮説 (H_0)

母集団の中央値はリッカート尺度における中央の値 3 と等しい．

対立仮説 (H_1)

母集団の中央値はリッカート尺度における中央の値 3 よりも大きい．

本検定法では，標本の各データと標本の代表値とリッカート尺度における中央の値 $\mu_0 = 3$ の差から統計量 T を求める．この T が Wilcoxon 符号順位検定数表 [14] における棄却限界値以下のとき帰無仮説を棄却する．すなわち，母集団の中央値は μ_0 よりも大きいと結論する．標本サイズが 25 を超える場合は T を用いて別の統計量を求めるが，本研究で扱う標本サイズは 25 以下のため上記の判断を行った．

4.1.2 実験結果

表 4, 5 は評価実験 A の各シナリオの結果を示している．表中の * がついている T は帰無仮説を棄却することを示している．

表 4 評価実験 A 表 2 の質問の回答結果

No.	シナリオ 1		シナリオ 2		シナリオ 3	
	棄却域	T	棄却域	T	棄却域	T
1	$T \leq 47$	*0.0	$T \leq 47$	*0.0	$T \leq 47$	*4.0
2	$T \leq 35$	36.0	$T \leq 35$	*24.5	$T \leq 35$	*13.5
3	$T \leq 47$	*6.0	$T \leq 53$	*0.0	$T \leq 53$	*19.5
4	$T \leq 47$	*0.0	$T \leq 53$	*4.5	$T \leq 53$	*10.0
5	$T \leq 60$	*0.0	$T \leq 41$	*0.0	$T \leq 41$	*10.0

表 5 評価実験 A 表 3 の質問の回答結果

No.	シナリオ 1		シナリオ 2		シナリオ 3	
	棄却域	T	棄却域	T	棄却域	T
1	$T \leq 35$	*30.0	$T \leq 47$	*18.0	$T \leq 47$	*26.0
2	$T \leq 30$	37.5	$T \leq 41$	*23.5	$T \leq 35$	*16.5
3	$T \leq 53$	*6.5	$T \leq 21$	24.0	$T \leq 30$	*11.0
4	$T \leq 47$	*2.0	$T \leq 60$	*4.5	$T \leq 53$	*0.0
5	$T \leq 47$	*6.5	$T \leq 53$	*3.5	$T \leq 47$	*4.5

表 4, 5 より，シナリオ 1 では「観点は分かりやすかったか」「推薦ゲームは気に入ったか」の質問項目でリッカート尺度の

中央の値 $\mu_0 = 3$ とは有意差が見られず、それ以外の質問項目では有意差が見られた。シナリオ 2 では「推薦されたゲームは見つけにくい良作だと感じたか」の質問項目で有意差が見られず、それ以外の質問項目で有意差が見られた。シナリオ 3 では全ての質問項目において有意差が見られた。

4.1.3 考察

シナリオ 1 では「観点は分かりやすかったか」「推薦されたゲームは気に入ったか」の 2 項目で有意差を示すことができなかった。被験者から感じたことや意見を自由記述として得られたので、その中からどのようなことが課題点となるのかを考察する。

「観点は分かりやすかったか」という質問項目と関連がある意見として

- 「多面的な没入体験」の「質の鑑賞」を中心に検索しましたが、ぱっと見たときに五感のどの部分と関連があるのかが分かりづらいかと感じました。
- 「物語への没入」について検索しようとしたが、「物語と没入体験」、「物語表現・没入要素」や「没入型物語」など同じような言葉の観点が多く、違いがわからなかった。

というものが得られた。これはレビューから抽出した観点に不適切なものが混ざってしまい、結果として分かりづらい名前になってしまったのだと考えられる。その原因として、Qwen3-8B に経験価値について十分な情報を持たせられなかった点や、同じような主張をしている観点の集約が不足している点が考えられる。

「推薦されたゲームは気に入ったか」に関しては

- 推薦されたゲームがどのようなものかぱっと見でわかるとさらに良いと感じた。例えば、ゲームジャンルなどは別枠で表示したりすると、どんなゲームかが想像しやすい。
- 自分が選んだゲームが動画が反映されていないゲームであったため、ゲームの操作感やシナリオにある実際の音声・サウンドを評価できないことに残念に思った。

など、推薦されたゲームの見せ方に関して否定的な意見が見られた。また

- 選択する観点が多いと感じた。そのため、選択すべき観点を探すことに少し手間取ってしまい、使いづらさを感じた。
- 観点が多く、探しにくい

など、観点での検索に関しても否定的な意見が見られた。

シナリオ 2 においても「観点は分かりやすかったか」「推薦されたゲームは気に入ったか」の 2 項目は有意差は見られたものの、棄却域と検定統計量 T との差が他の質問項目と比較して小さい結果になっている。

シナリオ 3 では「推薦されたゲームを実際にプレイしたいと感じたか」の質問項目が、棄却域と T との差が他の質問項目と比較して小さい結果になった。この項目に関して低く評価した被験者の 1 つの意見として

- 「ストーリーとキャラクター」という観点を選択したところ、レビュー文では批判的な意見での観点抽出が行われていたため、プラス面でのゲーム推薦が行われると考えていた自分としてはギャップに感じた。

という記述があった。実際に今回は感情極性を考慮せずに観点抽出を行ったので、ネガティブな意見がノイズになってしまったのだと考えられる。

以上の 3 つのシナリオの結果を総合すると、観点抽出におけるポジティブ・ネガティブのフィルタリング、経験価値を踏まえた SLM のプロンプティングやファインチューニング、検索・検索結果画面での大まかな情報と詳細情報の分別と、それらの効果的な見せ方をよく検討する必要がある。

4.2 評価実験 B

4.2.1 実験概要

一般的に UI はシステムを運用することで気づく不確定要素が多く、これには UI の重要性に気付くことが難しい点と関係している可能性がある [3]。そのため、評価実験 B では被験者は提案システムのインタフェースおよびそのユーザビリティを評価するため、表 6, 7 の質問に回答する。質問への回答は 5 段階リッカート尺度における評価である。

表 6, 7 の項目も、評価実験 A と同様に豊田ら [10] の研究に基づいて設定した。

表 6 評価実験 B UI の質問項目

No.	質問内容
1	文字の大きさは適切だったか
2	レイアウトは適切だったか
3	観点の選択や追加はしやすいと感じたか
4	選択観点エリアのサブグループのカテゴリは適切だったか
5	推薦結果一覧は、ゲームの概要を素早く把握できるスタイルで構成されていたか
6	詳細情報画面は、ゲーム推薦における重大な情報が表示されていたか

表 7 System Usability Scale (SUS) 質問項目

No.	質問内容
1	頻繁に使用したいと思う
2	不必要に複雑だった
3	使いやすいと感じた
4	利用するには技術者のサポートが必要だと思う
5	機能はまとまっていると思う
6	矛盾がとても多いと感じた
7	ほとんどの人がすぐに使いこなせると思う
8	使うのが面倒に感じる
9	自信をもって操作できた
10	使いこなすにはたくさんの事前知識が必要だと思う

表 6 の回答は、評価実験 A と同様に Wilcoxon 符号順位検定を用いて有意差検定を行った。表 7 の回答は、System Usability Scale (SUS) [1] を用いて評価を行った。SUS はそのシステムがユーザにとって使いやすいかどうかを評価するスコアであり、スコアが高いとユーザビリティが高い。平均の SUS スコアは 68 であり、この値を上回ると一般的に使いやすいシステムといえる。

4.2.2 実験結果

評価実験 B で行った UI 評価の結果は表 8 の通りである。表 8 より、すべての質問項目においてリッカート尺度における中央の値 $\mu_0 = 3$ との有意差が確認できた。

表 8 評価実験 B UI 評価の結果

質問内容	棄却域	T
文字の大きさは適切だったか	$T \leq 60$	*2.5
レイアウトは適切だったか	$T \leq 53$	*9.0
観点の選択や追加はしやすいと感じたか	$T \leq 41$	*20.0
選択観点エリアのサブグループのカテゴリは適切だったか	$T \leq 35$	*12.0
推薦結果一覧は、ゲームの概要を素早く把握できるスタイルで構成されていたか	$T \leq 60$	*12.0
詳細情報画面は、ゲーム推薦における重大な情報が表示されていたか	$T \leq 60$	*5.0

また、SUS の評価実験の結果を表 9 にまとめた。太字の値は SUS の平均といわれている 68 点よりもスコアが上回っていることを示す。表 9 より、各被験者において平均を上回っているスコアは見られたが、全体的なスコア（被験者平均）は平均を下回ったので、本システムは使いやすいとは言えない。

表 9 評価実験 B SUS の結果

被験者 No.	SUS スコア	被験者 No.	SUS スコア
1	70.0	11	75.0
2	50.0	12	47.5
3	70.0	13	35.0
4	90.0	14	87.5
5	67.5	15	82.5
6	87.5	16	60.0
7	40.0	17	52.5
8	50.0	18	25.0
9	80.0	19	67.5
10	77.5	20	45.0
被験者平均		63.0	

4.2.3 考察

リッカート尺度における中央の値 $\mu_0 = 3$ との有意差は示されたが、棄却域と T との差が小さい質問項目として

- 観点の選択や追加はしやすいと感じたか
- 選択観点エリアのサブグループのカテゴリは適切だったかが挙げられる。これらに関連する意見として
- 観点が多く、すべてそれなりの長さの日本語で表示されているので、読むのがおっくうでした
- ゲーム検索を行う際、少し項目がどこにあるのか分かりづらいように感じた
- 検索画面において、折り畳みボックスとその中身の整合性をより高めた方が良いと感じた

が見られた。これは、4.1 節と同様に観点抽出の段階で不適切な点があったことがうかがえる。UI の面では、観点の概要を素早く把握させられるレイアウトや、観点項目を分かりやすくするようなレイアウトを検討する必要があると考える。

ユーザビリティ面は特に UI やシステム機能に関する意見が多く見られた。例として

- 更新以外で一度チェックしたものを一括でリセットする機能があるといいなと思いました
- ヘッダーのログアウトの位置が検索画面とゲーム一覧画面で異なるので、ログアウトを一番右に移動するとわかりやすい

などが挙げられる。これらの意見に基づいてシステムを改善して評価テストを繰り返し行うことで、ユーザビリティが向上すると期待される。

5 おわりに

本研究では、SLM を用いてレビューから抽出した経験価値を利用したゲーム推薦システムを構築した。ユーザレビューデータや各ゲームの動画・紹介テキストなどのシステム実装に必要なデータは、Steam Web API のデータ取得 URL にアクセスして取得した。経験価値抽出は複数のレビュー集約プロセスによって構成されている。その中で観点をサブ/類似評価グループに集約するプロセスでは、観点リストを固定長に区切った小集合内でクラスタリングし、得られたグループ名リストをまた同様に処理するという、階層型クラスタリングに着想を得た独自の方法で処理を行った。以上の処理によって得られたデータを用いて Flask によって Web システムを実装した。このシステムを用いて被験者実験を行い、経験価値に基づく観点とシステムの有効性を評価した。

実験の結果、観点に関しては多くの項目で有効性を示すことができたが、主に観点の分かりやすさについて課題が残った。UI に関しても一定の効果を示すことができたが、観点の見せ方について課題が残った。ユーザビリティに関しては残念ながら SUS の平均スコアを上回ることができず、使いやすいシステムであることを示すことができなかった。

今後の課題としては、ポジティブな意見のレビューのみフィルタリング、SLM のプロンプティングの改善やファインチューニングが挙げられる。これらは経験価値抽出の精度向上に必要な改善点である。UI やユーザビリティは、被験者実験アンケートで得られた自由記述意見に基づいて改善していく必要がある。

また、本研究では対象領域をゲームに限定して分析を行ったが、本研究で扱った「経験価値」はマーケティング分野において広く用いられている概念であり、特定のコンテンツ領域に依存しない汎用性を有すると考えられる。このため、本研究で提案した枠組みは音楽や旅行などの他の体験型コンテンツにも応用可能であると期待される。対象領域の拡張を通じて、本手法の有効性および汎用性をさらに検証し、より汎用的かつスケラブルな分析・推薦手法へと発展させることが望まれる。

文献

- [1] J. Brooke: SUS: A Quick and Dirty Usability Scale. Usability Evaluation in Industry, Part 6, Chapter 21, pp. 189–194,

- 1996.
- [2] A. Gu and T. Dao: Mamba: Linear-Time Sequence Modeling with Selective State Spaces. arXiv preprint arXiv:2312.00752, 2024.
 - [3] 風間正弘, 飯塚洗二郎, 松村優也: 推薦システム実践入門 一仕事でつかえる導入ガイド. オライリー・ジャパン, 2022.
 - [4] 樹下雅章: Flask 本格入門 ~やさしくわかる Web アプリ開発~. 技術評論社, 2023.
 - [5] A. Korikov, G. Saad, E. Baron, M. Khan, M. Shah and S. Sanner: Multi-Aspect Reviewed-Item Retrieval via LLM Query Decomposition and Aspect Fusion. SIGIR'24 Workshop on Information Retrieval's Role in RAG Systems, 2024.
 - [6] 中井香那子, 山本岳洋, 大島裕明: 大規模言語モデルを用いた商品比較のためのレビュー集約. 第 17 回データ工学と情報マネジメントに関するフォーラム (DEIM 2025), 3A-03, 2025.
 - [7] 中谷知博, 星野准一: 経験価値の分類に基づくゲーム推薦システム. 社会法人 情報処理学会 研究報告, 2008-EC-11, Vol. 2008, No. 129, pp. 49-56, 2008.
 - [8] 大山浩暉, 竹川佳成, 平田圭二: レビュー文を考慮したゲーム推薦システムの実現に向けた単語の類似度調整の取り組み. エンタテインメントコンピューティングシンポジウム (EC2017), 2017.
 - [9] B. H. Schmitt: Experiential Marketing. Journal of Marketing Management, Vol. 15, No. 1, 1999.
 - [10] 豊田陽大, 上野史, 太田学: レビューから抽出した観点をを用いた飲食店推薦システムの提案. 第 17 回データ工学と情報マネジメントに関するフォーラム (DEIM 2025), 9E-03, 2025.
 - [11] H. Wu, Y. Yuan, L. Mikaelyan, A. Meulemans, X. Liu, J. Hensman and B. Mitra: Learning to Extract Structured Entities Using Language Models. The 2024 Conference on Empirical Methods in Natural Language Processing, pp. 6817-6834, 2024.
 - [12] 4Gamer.net: 【島国大和】ゲームと他のメディアは何がどうちがうんだー! というお話. <https://www.4gamer.net/games/000/G000000/20100604072/>.
 - [13] はてなブログ: SteamAPI に軽く触れた記憶. <https://271108.hatenablog.com/entry/2023/02/08/204104>.
 - [14] 統計検定のまとめブログ: ウィルコクソンの符号順位検定. https://data-science.gr.jp/theory/tst_wilcoxon_signed_rank_test.html.