属性に基づいた類似商品レビュー検索による評価情報補完インタフェース

藤井真梨乃† 山本 岳洋† 湯本 高行†

† 兵庫県立大学 大学院情報科学研究科 〒 651–2197 兵庫県神戸市西区学園西町 8–2–1
E-mail: †ad24f053@guh.u-hyogo.ac.jp, ††{t.yamamoto,yumoto}@sis.u-hyogo.ac.jp

あらまし 本研究では，レビューが少ない商品の評価を，類似商品のレビューを用いて補完するインタフェースを提案する．例えば，ユーザがある新商品のテレビについて「どれぐらい場所を取るのか」を調べるときに，大きさが類似する商品に存在する大きさに関するレビューを提示する．これにより，ユーザは商品購入の判断材料を増やすことができる．この手法の実現に向け，我々はこれまでに，入力に関連する属性の判定や属性に関連する意見の抽出に LLM を用いる手法を提案してきた．本研究では，この手法を用いて評価情報を補完するインタフェースを実装し，ユーザ実験を通じてその有効性を検証する．具体的には，レビューが存在しない場合に類似商品から意見を補完するインタフェースとそうでないインタフェースを用いて実験を行い，類似商品の意見を表示することがユーザの意思決定に有用であるかを検証した．参加者が回答したアンケートを分析した結果，補完を行わないものと比べて補完を行うインタフェースの有用性は高いことが示された．また，類似商品やその意見への信頼度は中立的であり，補完された意見に異なる情報も含まれている前提でユーザは検索を行うことが示された．

キーワード インタフェース，レビュー検索，ユーザ支援

1 はじめに

EC サイトで商品を購入する上で，レビューは購入の判断材料となる．インターネットの発展に伴ってオンラインで買い物をする人が増加しており，経済産業省の調査によると，物販系分野の BtoC-EC 市場規模は 2013 年の 5 兆 9,931 億円から 2022 年の 13 兆 9,997 億円まで年々拡大している [19].

しかし，新商品やマイナーな商品など，レビューが少ない商品では，ユーザは知りたい情報を満足に得られないという問題がある．例えば A という新製品のテレビについて，画面の大きさに関するレビューを調べたかったとする．このとき，画面の大きさに関するレビューを検索しても，それについて述べられているレビューが少ない，もしくは存在しないことがある．

そこで, この問題を解決するために, 大規模言語モデル（LLM）を用いてレビューを補完する手法を提案する．まずユーザは調べたい商品を選択する．次に,「どれぐらい場所を取るか」といった調べたいことを文章で入力し，その文章から LLM を用いて商品説明のスペック情報に存在する「サイズ」といった属性を選択する．そして，選択した属性の値が類似する商品から,「6 畳和室に置いていますが，ジャストサイズ！」といった属性について述べているレビューを LLM を用いて取得し，ユーザに返答するという手法である．これにより，ユーザは商品についているレビューが少なくても，商品に関する評価をある程度予測することができる．例えば，テレビ A と同じ画面サイズを持つテレビ B に「10 畳の部屋にぴったりな大きさだと思います!」といったレビューが付いていた場合，これはテレビ A でも同様であるといえる．

このような，類似する商品のレビューを用いて検索している商品の評価を補完する研究は以前から行われている [3] [11] が，LLM を用いたものは存在しない．LLM を用いることによって，ユーザによる入力の自由度や処理精度が向上すると考えた．

我々はこれまでに，この手法中の LLM の判定精度や，この手法によって取得した意見がどの程度正しいのかを検証してきた [6]．しかし，実際に手法を適用したインタフェースの有用性についての検証は行っていない．本研究では，この手法を用いて実際に評価情報を補完するインタフェースを実装し，ユーザ実験により手法の有効性を検証した．補完を行わないインタフェースを比較手法として用意し，両方使用してもらった上でアンケートに回答してもらった．その結果，ある商品に商品レビューが存在しないとき，比較手法である補完を行わないものよりも補完を行うインタフェースの有用性が高いことが示された．また，類似商品およびその意見への信頼度は高いとも低いともいえず，ユーザは補完された意見に異なる情報も含まれている前提で検索を行っているという結果になった．

2 関連研究

2.1 商品レビュー分析

レビューから属性または観点とその意見を抽出する研究が行われている．Hu らは，商品の属性に対して，肯定的，否定的な意見を WordNet を用いて収集した [8]．杉木らは，自然言語表現で入力されたクエリと商品レビューに対し，感性表現シソーラスを用いて（項目，値）の組を抽出した [17]．Poria らは，埋め込み表現を用いてレビュー中で述べられている観点を抽出した [12]．Ishii らは，映画のレビューが何の観点について述べているのかを，T5 を用いて 1 文ごとに生成した [9]．Adhikary らは，LLM を用いてレビューの観点とその意見の抽出を行っ

た [1].

これらの研究では，商品レビューから属性や観点を抽出するという部分に重きを置いている．本研究では，LLM によって商品レビューから抽出された属性と意見を用いて，類似する商品の評価を補完する．

2.2 商品レビュー検索インタフェース

商品レビューの検索を行うインタフェースについての研究が行われている．平山らは，商品レビューに形態素解析を用いて語の共起関係を抽出し，評価属性ごとの記述および関係性を可視化したインタフェースを作成した [16]．市村は，係り受け解析を用いて飲食店のレビューから料理に関する文の要約を 3 種類行い，インタフェースを作成して提示したところ，料理に関する抽出を行った 1 文をそのまま載せたシステムが最も満足度が高いことを報告した [15]．Jasim らは，肯定的，中立，否定的なレビューの訪問率を可視化してバイアスを減らすインタフェースを提案した [10]．また，我々は以前，ユーザが入力したクエリを基に LLM を用いて要約されたレビューを提示するシステムを提案し，結果の見やすさ等がレビューを要約せずに表示するときより向上することを確認した [5]．

これらの研究では，レビューが多い商品の検索の支援を目的としている．本研究では，レビューが少ない商品に対して，類似している別の商品のレビューを補完することの有効性を検証している．

2.3 類似商品のレビューから評価を取得する研究

類似商品のレビューから評価を取得する手法の研究が行われている．Park らは，クエリ尤度モデルを用いて，調べたい属性が類似している商品から属性に関するレビューを取得した [11]．Du らは，有用度や網羅性も考慮して，調べたい属性が類似している商品から属性に関するレビューを取得した [3]．Rozen らは，ある商品のレビューには回答が書いていない質問に対して，RoBERTa で類似商品と質問を考慮できるモデルを構築し，別の商品の回答を基に補完する手法を提案した [13]．

本研究では，類似商品から評価を取得するときに LLM を用いている．また，以上の研究では行われていない，類似商品から評価を取得して実際にレビュー検索インタフェースに表示させる場合のユーザ評価を行っている．

3 類似商品レビューを用いた評価情報補完インタフェース

本節ではまず，本研究で実現したインタフェースの概要について述べる．次に，このインタフェースで使用している，我々がこれまでに提案した類似商品からのレビュー補完手法 [6] について述べる．最後に，インタフェースの詳細な操作方法について説明する．

3.1 インタフェースの概要

インタフェースの概要を図 1 に示す．まず，ユーザは商品一覧から調べたい商品を選択した後，調べたいトピックを選択する．このとき，調べたい商品を**商品クエリ**，調べたいトピックを**テキストクエリ**と定義する．図の例では，ユーザが「TID920」というドライヤーを調べており，「風量の強さ」に関してレビューを検索しようとしている．このとき，「TID920」が商品クエリ，「風量の強さ」がテキストクエリとなる．検索ボタンをクリックすると，テキストクエリに関連する属性が類似している商品から，属性に関する意見を表示する．テキストクエリに関連する属性を**属性クエリ**，属性クエリが類似する商品を**類似商品**と定義する．図の例では，「風量の強さ」というテキストクエリを入力すると，「風量」という属性クエリが類似している「TID320」という類似商品から，「風力があり」という意見が補完されている．

3.2 類似商品からのレビュー補完の概要

3.1 節で述べたインタフェースで使用している，類似商品からレビューを補完する手法 [6] の概要について述べる．この手法は大きく 3 つの段階に分かれている．

1. テキストクエリと関連する属性を，LLM を用いて判定する．
2. 属性クエリが類似する商品を選択する．
3. 類似商品のレビューから属性クエリに関する記述を LLM を用いて抽出し，ランキングする．

次節より，この 3 段階について詳しく述べる．

3.3 属性クエリ判定

テキストクエリと関連する属性を，LLM を用いて判定する．本研究では，テキストクエリを商品クエリの機能に関する文章と限定する．機能に限定する理由は，補完するために用いる商品属性には機能に関するものが多いためである．値段やデザイン（カラー等）もレビューを読む上で重要な要素であるが，購入するタイミングによって変動するため，本研究では除外する．テレビについて検索を行う際のプロンプトは以下の通りである．

> あなたは今{商品クエリのジャンル}のレビューを検索しようとしています．
> 「{テキストクエリ}」という入力に対して，「{属性クエリの候補}」という属性はあてはまりますか．ステップバイステップで考え，出力としてはあてはまるかを True/False で書いてください．
> ###OUTPUT_EXAMPLE:
> True or False

このとき，{商品クエリのジャンル}には「テレビ」，{テキストクエリ}には「どれぐらい場所を取るか」や「画質の良さ」，{属性クエリの候補}として「画面サイズ」や「画素数」などが入力される．このプロンプトで True と出力された属性を属性クエリとして利用する．本研究では，この判定を gpt-4o-2024-0806[1] を使用して行っている．

1：https://platform.openai.com/docs/models\#gpt-4o



図 1　インタフェースの概要.

3.4　類似商品の選択

商品クエリにおける属性クエリの値と，それ以外の商品における属性クエリの値の間で類似度を計算し，類似商品を選択する．このとき，3.3 節で True と判定された属性のみを属性クエリ集合 $A_q^{(i)}$ として使用する．ある商品 $P^{(i)}$ の属性 $p_l^{(i)} = (a_l^{(i)}, v_l^{(i)})$ について，商品 $P^{(j)}$ の対応する属性 $p_l^{(j)} = (a_l^{(j)}, v_l^{(j)})$ との類似度を求める．まず，属性が量的データの場合，以下の式に従って類似度を求める．

$$\mathrm{sim}(p_l^{(i)}, p_l^{(j)}) = 1 - \left| v_l^{(i)} - v_l^{(j)} \right| \tag{1}$$

ここで，$v_x^{(i)}$ は最大値が 1 となるように正規化された値である．また，属性がカテゴリカルデータである場合，以下の式に従って属性値が完全に一致する場合にのみ類似度に 1 を加算する．

$$\mathrm{sim}(p_l^{(i)}, p_l^{(j)}) = \begin{cases} 1 & \text{if } v_l^{(i)} = v_l^{(j)} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

本研究では，属性クエリ集合 $A_q^{(i)}$ の各属性についてこれらの式で類似度を求め，スコアの高い商品から上位 10 件の商品を類似商品として用いる．類似度が同値だった場合はレビュー数が多い商品を上位とする．

3.5　商品レビューからの意見の抽出，ランキング

商品レビューの中で属性について述べている部分を LLM を用いて抽出する．属性について述べている部分を抽出し，述べていない属性については「なし」と出力することで，関連があるかを判定することができる．

例えば，テレビのレビュー「6 畳の部屋で大きすぎず見やすく買ってよかったです．設置もとても簡単でした． 発送もとてもはやくて助かりました．ありがとうございました．」を用いる場合，プロンプトは以下の通りである．

> テレビのレビュー「6 畳の部屋で大きすぎず見やすく買ってよかったです．設置もとても簡単でした． 発送もとてもはやくて助かりました．ありがとうございました．」から「{属性クエリの 1 つ}」に関連する記述を抜き出してください．OUTPUT_EXAMPLE の形式に従って出力してください．
> 関連する記述が無い場合，attribute の値は「{属性クエリの 1 つ}」，sentence の値は「なし」と出力してください．
> {json で出力するための指示}
> ###INPUT_EXAMPLE:
> （ドライヤーのレビューの場合）
> このドライヤー，パワーが強くてすぐ乾きます．
> ###OUTPUT_EXAMPLE:
> （ドライヤーのレビューから風量に関連する記述を抜き出す場合）
> {{attribute: 風量, sentence: パワーが強くてすぐ乾く }}
> INPUT:6 畳の部屋で大きすぎず見やすく買ってよかったです．設置もとても簡単でした． 発送もとてもはやくて助かりました．ありがとうございました．
> JSON_OUTPUT:

精度の高い回答を生成するために，few-shot プロンプトを用いて例を与えた．このとき，例が出力に影響を与えないようにするため，現在調べているジャンルとは別ジャンルの属性とレビューを使用した．この例では，「テレビ」とは別のジャンルである「ドライヤー」というジャンルの例を与えている．

属性クエリとして「画面サイズ」を入力すると，画面サイズに関連する意見である「デカくて」という記述が出力される．一方，属性クエリとして「録画機能」を入力してプロンプトを実行すると，レビューに関連する記述が存在しないため，出力

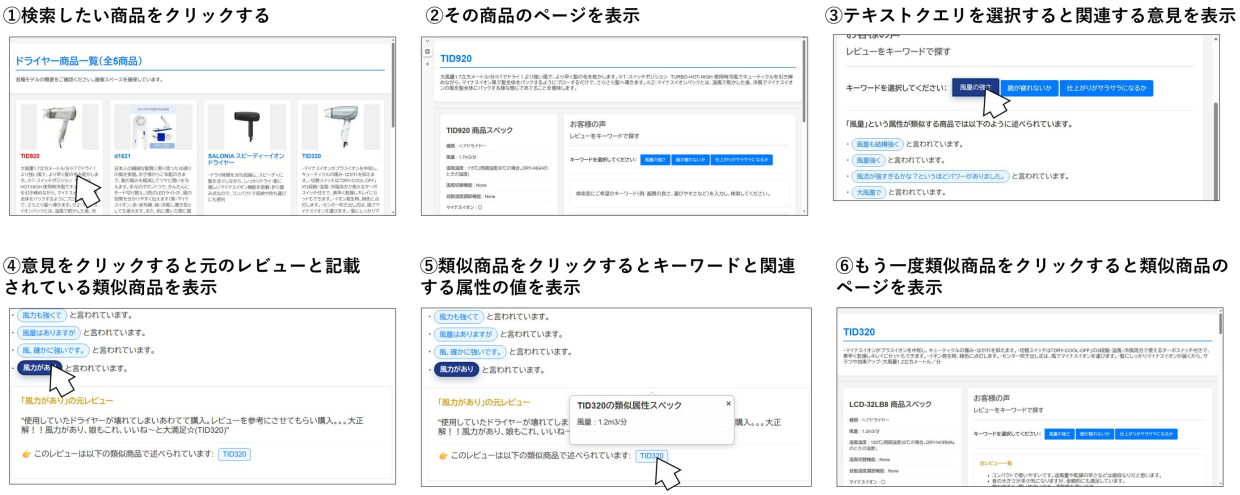


図 2　インタフェースの操作方法.

は「なし」となる.

本研究では，`Llama-3.1-Swallow-8B-Instruct-v0.3` [4] [7] を使用して以上の操作を行っている．このモデルは Llama3.1 から継続事前学習を行い，日本語能力を強化したモデルである.

類似商品のレビュー中に存在する意見を取得した後，それらをランキングしてユーザに提示する．意見をランキングするために，テキストクエリとこれらの意見の類似度を計算した．具体的には，テキストクエリと類似商品の意見をベクトル化し，コサイン類似度を求める．本研究では，テキストクエリと意見をベクトル化する際に multilingual-e5-large[2] [14] を用いている.

3.6　評価情報補完インタフェース

評価情報を補完するインタフェースの操作方法を図 2 に示す.

まず，初期画面として商品一覧ページが表示される．そして，検索したい商品（商品クエリ）をクリックすると，その商品のページが表示される．このページには，商品説明文，スペック情報，テキストクエリが記載されている．提案手法では，ユーザはテキストクエリを自由に入力することが可能である．しかし，実験では対応可能なクエリを絞るために，今回実装したインタフェースではテキストクエリをあらかじめ候補として示し，ユーザが選択できるようにしている．1 つのテキストクエリを選択すると，それに関連する意見が表示される．このとき，商品クエリのレビューを擬似的に 0 件として，類似商品から取得した意見を表示している．表示する意見を取得する際に，前節までで述べた手法 [6] を用いている．意見をクリックすると，意見の抽出元のレビューと，そのレビューがついている類似商品が表示される．類似商品をクリックすると，テキストクエリと関連する属性の値が表示される．また類似商品を再度クリックすると，その商品の商品ページが表示される．類似商品のページでは，その商品についている全てのレビューと，テキストクエリに関連する意見が表示できる．これにより，類似商品のページに行って，商品クエリのページで見たレビューが存在することを確認できるようになっている.

4　実　　験

4.1　データセット

はじめに，本研究で使用するデータセットについて説明する．実験は，テレビとカメラの 2 つのジャンルで行った．この 2 つのジャンルを選んだ理由は，いずれのジャンルも属性数が多いためである.

属性は，価格.com[3] 上に掲載されているテレビ及びカメラジャンルのスペック情報を使用した．価格.com とは，商品説明やそれに対するレビューなどを掲載している日本のウェブサイトである．商品説明の中にスペック情報という属性情報をまとめたページがあり，ジャンルが同じであればどの商品にも同じ属性が存在している．テレビは 22 個，カメラは 19 個の属性を使用した．使用した属性を表 1 に示す.

商品レビューについては楽天データセットに含まれる楽天市場の商品レビューデータ [18] を利用した．具体的には，2017 年 1 月 1 日〜2019 年 12 月 31 日に投稿されたテレビとカメラジャンルのレビューの中から，レビュー数上位 100 商品を抜き出した．そのうち，型番が明記されており，かつ価格.com 上にスペック情報が掲載されている商品とレビューを使用した．テレビジャンルでは 87 商品 7,963 件のレビュー，カメラジャンルでは 55 商品 1,840 件のレビューを使用し，テキストクエリと類似する 10 件の意見を取り出した.

テキストクエリについては，著者らで各ジャンル 7 個ずつ用意した．用意したクエリを表 2 に示す.

4.2　実験参加者

実験参加者は，兵庫県立大学に所属する学生 12 名である．2025 年 12 月 23 日から 2026 年 1 月 5 日にかけて実験を行い，参加者には約 2,000 円を支払った.

4.3　比 較 手 法

比較手法は，類似商品からレビューを補完しないインタフェー

2：`https://huggingface.co/intfloat/multilingual-e5-large`

3：`https://kakaku.com/`

表 1　カメラとテレビの属性一覧.

ジャンル	属性
テレビ	画面サイズ, 画素数, パネル種類, HDR 方式, 映像処理エンジン, バックライト, 量子ドット, 倍速機能, BS 8K, BS 4K/110 度 CS 4K, 録画機能, ドライブ内蔵, 自動録画機能, 2 番組同時録画, 早見再生, スマートスピーカー連携, 回転式スタンド, HDMI 端子, リモコン（音声操作）, スピーカー数, 幅 x 高さ x 奥行, 重量
カメラ	セルフタイマー, 自分撮り機能, 液晶モニター, 起動時間, 撮影枚数, シャッタースピード, 撮影感度, 重量, 幅 x 高さ x 奥行き, インターフェース, タイムラプス, 画素数, 連写撮影, 手ブレ補正機構, 4K 対応, タッチパネル, 記録メディア, 記録フォーマット, 撮像素子

表 2　用意したテキストクエリ.

テレビ	カメラ
運びやすさ	野球の静止画を撮るなら
録画設定はしやすいか	液晶画面の見やすさ
音質の良さ	振動耐性について（手ぶれ）
脚の安定性	電池持ちの良さ
画質の良さ	スマホよりサッと撮れるか
コンサートを観るのに向いているか	容量の大きさ
スポーツ観戦に向いているか	星空の撮影は可能か

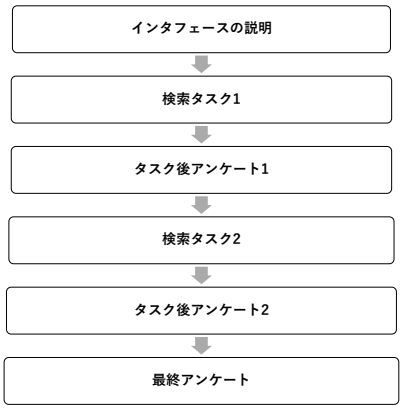


図 3　実験手順.

スである．このインタフェースでは，商品クエリのページにおいて 1 つテキストクエリを選択したとき，「{テキストクエリ}というキーワードについて，この商品で述べられているものはありません.」という文言のみを表示し，類似商品による補完がない状態になる．それ以外の仕様は 3.6 節で述べたインタフェースと同一であり，商品クエリ以外の商品では，その商品についている全てのレビューと，テキストクエリに関連する意見が表示できる．

4.4　実 験 手 順

実験手順を図 3 に示す．まず，実験参加者に対して本実験で用いるインタフェースの概要について説明を行った．その後，操作に慣れてもらうために，インタフェースを用いてドライヤー 1 商品についているレビューの検索を自由に行ってもらった．次に検索タスクを行った．具体的には，テレビ 1 商品（40FB10P），カメラ 1 商品（COOLPIX A100）について，どのような商品であるのかをレビューを参照しながら，検索結果を記録してもらうというものである．本実験は被験者内実験として行っており，各参加者に検索してもらった商品とインタフェースの順番は表 3 の通りである．1 商品の検索タスクが終わるごとに，タスク後アンケートに回答してもらった．タスク後アンケートでは，使いやすさや商品への理解度などを回答してもらった．全ての商品のタスク後アンケートが終わった後，最終アンケートに回答してもらい，ユーザ実験を終了した．最終アンケートでは，各インタフェースの良かった点と悪かった点や，補完された商品の妥当性などの提案手法に限定した設問について回答してもらった．

表 3　各参加者が検索を行う順番.

参加者	検索タスク 1	検索タスク 2
1, 2, 3	提案手法/テレビ	比較手法/カメラ
4, 5, 6	比較手法/テレビ	提案手法/カメラ
7, 8, 9	提案手法/カメラ	比較手法/テレビ
10, 11, 12	比較手法/カメラ	提案手法/テレビ

4.5　検索タスク

検索タスクでは，実験参加者にある 1 商品について複数の観点から調べ，文章にまとめるというタスクを行ってもらった．実験参加者は以下の配布された資料から想定される状況を把握したのちに，検索タスクを開始した．

> あなたは{テレビ or カメラ}を買い替えようとしており，友人の勧めで「{40FB10P or COOLPIX A100}」という商品に興味を持っています．そこで，その商品について以下の観点で調べてみることにしました．
> （表 2 参照）
> 他の商品のレビューも見ながら，この{テレビ or カメラ}がどのような商品なのかをメモにまとめてください．

検索する観点については，表 2 のテキストクエリを使用した．制限時間は 1 商品につき 15 分とした．理由は，7 つの観点で検索するにあたり，各観点の検索時間として 2 分程度が適切であると判断したためである．また，参加者には満足するまで調べ終えたと感じたら検索を切り上げても良いと伝えた．

本タスクでは，回答の単語数を測定することにより，RQ1 について検証した．理由は，回答の単語数が多ければ回答の具

表 4　タスク後アンケートの項目.

評価項目	質問
意見の網羅性	同じ観点の中で漏れなく意見を調べることができたと思う.
商品への理解度	このシステムによって，調べたい商品について理解できたと思う.
レビュー抽出の精度	検索結果に提示された意見は，元のレビューの主旨と相違ないと思う.
全体的な満足度	システムの機能に満足している.

体性が高く，意見の網羅性や商品への理解度を調べることが可能であると考えたからである．具体的には，回答中のストップワードを処理し，テキストクエリ名ごとに箇条書きを行っている場合（例：「運びやすさ：良い」など）にはそのテキストクエリの部分を削除し，単語数をカウントした．

4.6　アンケート

本実験ではタスク後アンケート，最終アンケートの 2 種類のアンケートを行った．タスク後アンケートでは，それぞれのタスクにつき 14 個の設問を用意した．そのうち 10 問は SUS(System Usability Scale) [2] の設問である．SUS とはシステムの利便性を評価するための尺度であり，0 から 100 の範囲で算出される指標である．SUS を用いて評価されたさまざまなシステムの平均は 68 点と言われている．残りの 4 問の項目は表 4 の通りである．これらの設問は，主に各インタフェースの有用性と，それを用いて行った検索の満足度について，すなわち RQ1 について検証するために用意したものである．

最終アンケートでは，各インタフェースの良かった点と悪かった点を自由記述で書いてもらった後，提案手法のインタフェースに限定した設問を 3 問用意した．この 3 問の項目は表 5 の通りである．これらの設問は，類似商品提示の有用性については RQ1，それ以外の質問は RQ2 を検証するために用意したものである．

本実験では，自由記述以外の全ての設問で 5 段階リッカート尺度を用いた（1：そう思わない，2：あまりそう思わない，3：どちらともいえない，4：ややそう思う，5：そう思う）．

4.7　結　　果

ユーザ実験の結果の平均を表 6 に示す．検索タスクおよびタスク後アンケートのすべての評価項目において，補完を行わない比較手法よりも補完を行う提案手法のインタフェースの方が上回るという結果になった．これらの項目において有意水準を 5%とした対応のある 2 標本 t 検定を行ったところ，回答の単語数およびレビュー抽出の精度以外の項目では有意差が認められた．加えて，最終アンケートでは類似商品提示の有用性が高いという結果になった．しかし，類似商品の精度と補完された意見の妥当性については値の平均が 5 段階リッカート尺度の中央値である 3 に近いという結果になった．また，各参加者の回答単語数の分布を図 4 に示す．実際に提案手法を用いたインタフェースで得られた回答の情報量が，比較手法と比べて増加している参加者が多いという結果になった．

5　考　　察

4.7 節で述べた，ユーザ実験の結果について考察する．

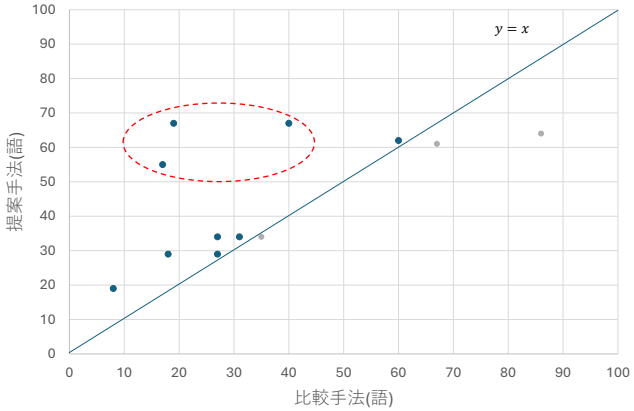


図 4　各参加者の回答の単語数分布.

5.1　RQ1: 類似商品のレビューを基に意見を補完することは，ユーザがレビュー検索を行う上で有用か

表 6 の検索タスクおよびタスク後アンケートの全評価項目において比較手法を上回ったこと，および最終アンケートにおいて類似商品提示の有用性の平均値が非常に高いことから，類似商品から意見を補完するインタフェースが補完しないインタフェースより有用であることが分かった．レビュー抽出の精度に有意差が認められなかったが，両手法において同じ要約方法を用いているため，自然な現象であると考えられる．また回答の単語数についても有意差が認められなかったが，提案手法においては補完された意見をそのまま書くことが多かったことに対し，比較手法では自力で探した類似商品についている様々な意見をまとめる必要があるため，比較手法での記述量が多くなる参加者が一部存在したことが原因であると考えられる．

有用である理由について，アンケートの自由記述を基に分析を行う．参加者の記述をまとめると，主に 3 つの観点から便利であると述べられていた．

1 つ目は，類似商品を簡単に探すことができる点である．参加者はこの点について以下のように述べている．

“類似商品のレビューが出てくる分，類似商品を探す手間が省けるので楽になった点が良かった．”（参加者 4/提案手法）

“類似商品をスペックから探してレビューを比較する必要があり大変だった”（参加者 9/比較手法）

比較手法では類似商品を自力で探し，レビューを比較する必要がある．一方，提案手法では類似商品を自動的に提示するため，ユーザは探す手間を省略することができたと考えられる．

2 つ目は，どの属性に着目すればよいかが分かる点である．参加者はこれについて以下のように述べている．

“良かった部分に関しては，それぞれの観点についてスペックのどの部分を見れば分かるのかを知らなかったため，スペックに関する知識がなくても観点について情報を得られた点です．”（参加者 7/提案手法）

“良くなかった点は，対象商品のレビューがなかったため，ほ

表 5　最終アンケートの項目.

評価項目	質問
類似商品の精度	検索結果に提示された類似している商品は，実際に類似していると思う.
類似商品提示の有用性	類似している商品を提示することは，検索を行う上で役に立ったと思う.
類似商品からの意見の妥当性	検索結果に提示された類似している商品の意見は，検索中の商品についていても自然であると思う.

表 6　ユーザ実験の結果．*は有意差が認められた値であることを示す（$p < 0.05$）.

		提案手法	比較手法
検索タスク	回答の単語数	46.3	36.3
タスク後アンケート	SUS	68.5*	57.5
	意見の網羅性	3.08*	1.50
	商品の理解度	3.25*	1.92
	レビュー抽出の精度	3.42	2.83
	全体的な満足度	3.58*	2.08
最終アンケート	類似商品の精度	3.25	–
	類似商品提示の有用性	4.33	–
	補完された意見の妥当性	3.17	–

かの商品のレビューとスペックを比較しながら商品の特徴を推察する必要があったため，事前知識がない場合は商品の特徴を捉えにくいと考えられる．”（参加者 11/比較手法）

レビューが無い商品のページでは，スペック情報と商品説明文のみしか存在しない状態である．特に事前知識を要するカメラなどのジャンルでは，比較手法のように関連する属性の提示が無い場合，テキストクエリに関連する商品属性を自力で特定することは困難である．図 4 に示した検索タスクにおける回答の単語数の分布を見ると，提案手法に比べて比較手法での単語数が著しく少ない参加者が存在する．これは，比較手法での検索において属性クエリを探すことが難しく，類似商品から得られる情報が限定的であったことを意味する．したがって，テキストクエリと関連する属性クエリを LLM を用いて選択して提示することにより，ユーザは着目すべき属性を簡単に見つけることが可能になったと考えられる．

3 つ目は，類似商品の意見から商品の使用感をある程度想像することが可能である点である．参加者はこの点について以下のように述べている．

“類似商品のレビューを参考に大きさなどのイメージだけはつかむことができた．”（参加者 2/提案手法）

“スペックだけでは分からない使用シーンの適応が想像しやすかった．”（参加者 5/提案手法）

レビューが無い商品では，カメラにおけるスポーツでの写真の撮りやすさといった，スペック情報だけでは分からない部分について知ることが難しい．そのため，類似商品の意見を提示することにより，その部分について推察することが可能になったと考えられる．

5.2　RQ2: ユーザは類似商品のレビューを基に補完された意見をどれぐらい信頼しているのか

表 6 の最終アンケートより，類似商品の精度と補完された意見の妥当性の値の平均が 5 段階リッカート尺度の中央値である 3 に近く，高いとも低いともいえないという結果になった．このような結果になった理由として，類似商品についている意見を参考程度に捉えており，完全には信頼していないことが挙げられる．参加者はこれについて以下のように述べている．

“テレビの画素数が結構差があったと感じたけど類似商品は本当に類似しているといえるのか”（参加者 3/提案手法）

“また，類似商品の意見を検索対象の商品に適用できない場合もあると思った．例えば，類似商品で，水中で撮影しているときに画面の白いモヤがあって見にくいとか．”（参加者 6/提案手法）

これらの意見から，参加者は類似商品から補完された意見が実際の商品でも完全に通用するとは考えておらず，異なる情報も含まれているという前提で検索を行っていたことが分かる．原因としては，LLM の誤判定により，類似商品についている意見への信頼度が下がっていることが考えられる．一方で，LLM の精度を改善することにも限界が存在するため，類似商品からの意見を完全には信頼せずに検索を行うというユーザの態度自体は妥当なものであると考えられる．

5.3　限　界　点

研究の限界点を 2 点述べる．1 点目は，本研究では 2 つのジャンルでしか評価を行っていない点である．そのため，提案手法が他のジャンルについても同様に有効なのかは明らかではない．例えば，ソファなどの家具では，サイズが同じでも素材が異なることにより使い心地について全く参考にならない場合が多いのではないかと考えられる．今後はより多様なジャンルを対象に評価を行う必要がある．

2 点目は，インタフェースの評価において，比較手法の利便性が著しく低く，適切に比較ができていない可能性がある点である．実際に表 6 を参照すると，SUS の点数の差が大きくなっており，意見の補完以外の部分において利便性に差が生じている可能性が示唆される．一般的な EC サイトの検索画面では，商品のサイズ等による絞り込み機能が備わっているが，本実験のインタフェースには備わっていない．提案手法ではレビューが存在しない商品でも類似商品が自動的に表示されるため，この機能が未実装であっても，多大な労力を払うことなく類似商品を見つけることが可能であった．一方で，比較手法では商品一覧ページから各商品のページに行って，商品の属性値を比較す

る必要があったため，類似商品を探すコストが非常に高くなっている．実際に比較手法の良くなかった点として，参加者 10 は「調べるのにものすごく時間がかかる」，参加者 4 は「商品ページで各属性を見つけるのが大変だったので，商品ページ内で見たい属性を選択するまたは一覧のページで商品属性を絞って検索できるといいなと思った.」と述べている．そのため，この機能を両手法の検索インタフェースに実装することで，より正確な有用性の比較が可能であったと考えられる．

6 ま と め

本研究では，レビューが少ない商品に対して，類似する商品につけられたレビューを用いてレビューを補完するという手法を適用したインタフェースの評価を行った．具体的には，まずユーザが商品と商品について調べたい文章を入力し，LLM を用いて文章から関連する属性を選択する．そして，関連する属性から類似商品を探し，LLM を用いて類似商品のレビューから関連する属性についての記述を取得するという手法である．ユーザ実験を行った結果，ある商品に商品レビューが存在しないとき，比較手法である補完を行わないインタフェースよりも補完を行うインタフェースの有用性が高いことが示された．また，類似商品の意見への信頼度は高いとも低いともいえず，ユーザは補完された意見に異なる情報も含まれているという前提で検索を行っているという結果になった．これらの結果から，類似商品のレビューを用いた意見の補完は，レビューの不足を補うインタフェースとしての有用性を備えていることが示唆された．

謝　　辞

本研究は，JSPS 科研費 JP24K03228 の助成を受けたものです．また，本研究では国立情報学研究所の IDR データセット提供サービスにより楽天グループ株式会社から提供を受けた「楽天データセット」(https://rit.rakuten.com/data_release/) を利用しました．ここに記して謝意を表します．

文　　献

[1] Jiban Adhikary, Mohammad Alqudah, and Arun Palghat Udayashankar. RevieWeaver: Weaving together review insights by leveraging LLMs and semantic similarity. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 3: Industry Track)*, pp. 431–448, 2025.

[2] John Brooke. SUS-a quick and dirty usability scale. *Usability evaluation in industry*, Vol. 189, No. 194, pp. 4–7, 1996.

[3] Na-na Du, Liu-tong Xu, and Hai Huang. MSRM: A novel model to retrieve meaningful opinion sentences for new products. In *Proceedings of the 2017 International Conference on Mathematics, Modelling and Simulation Technologies and Applications*, pp. 214–221, 2017.

[4] Kazuki Fujii, Taishi Nakamura, Mengsay Loem, Hiroki Iida, Masanari Ohi, Kakeru Hattori, Hirai Shota, Sakae Mizuki, Rio Yokota, and Naoaki Okazaki. Continual pre-training for cross-lingual LLM adaptation: Enhancing japanese language capabilities. In *Proceedings of the First Conference on Language Modeling*, 2024.

[5] Marino Fujii, Yuka Kawada, Takehiro Yamamoto, and Takayuki Yumoto. Review search interface based on search result summarization using large language model. In *The DASFAA 2024 Workshop on Emerging Results in Data Science and Engineering*, pp. 293–301, 2024.

[6] Marino Fujii, Takehiro Yamamoto, and Takayuki Yumoto. Supplementing product reviews: Retrieving opinions from products with similar attributes. In *Proceedings of the 27th International Conference on Information Integration and Web Intelligence*, pp. 15–30, 2025.

[7] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, and et al. The Llama 3 herd of models, 2024.

[8] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168–177, 2004.

[9] Tomohiro Ishii, Yoshiyuki Shoji, Takehiro Yamamoto, Hiroaki Ohshima, Sumio Fujita, and Martin J. Dürst. Generating fine-grained aspect names from movie review sentences using generative language model. In *Proceedings of the 25th International Conference on Information Integration and Web Intelligence*, pp. 217–232, 2023.

[10] Mahmood Jasim, Christopher Collins, Ali Sarvghad, and Narges Mahyar. Supporting serendipitous discovery and balanced analysis of online product reviews with interaction-driven metrics and bias-mitigating suggestions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–24, 2022.

[11] Dae Hoon Park, Hyun Duk Kim, ChengXiang Zhai, and Lifan Guo. Retrieval of relevant opinion sentences for new products. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 393–402, 2015.

[12] Soujanya Poria, Erik Cambria, and Alexander Gelbukh. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, Vol. 108, pp. 42–49, 2016.

[13] Ohad Rozen, David Carmel, Avihai Mejer, Vitaly Mirkis, and Yftah Ziser. Answering product-questions by utilizing questions from other contextually similar products. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 242–253, 2021.

[14] Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. Multilingual E5 text embeddings: A technical report. *arXiv preprint arXiv:2402.05672*, 2024.

[15] 市村哲. 口コミから美味しい料理店を手早く探すシステム. 情報処理学会論文誌, Vol. 61, No. 11, pp. 1748–1756, 2020.

[16] 平山拓央, 湯本高行, 新居学, 佐藤邦弘. 語の共起と極性に基づく商品レビュー閲覧支援システム. 情報処理学会研究報告, Vol. 2012–DBS–155, No. 3, pp. 1–9, 2012.

[17] 杉木健二, 松原茂樹. カスタマーレビューに基づく商品検索のための感性表現シソーラスの構築. 言語処理学会第 15 回年次大会発表論文集, pp. 781–784, 2009.

[18] 楽天グループ株式会社. 楽天市場データ. 国立情報学研究所情報学研究データリポジトリ. (データセット), 2020. https://doi.org/10.32130/idr.2.1.

[19] 経済産業省. 令和 4 年度 電子商取引に関する市場調査 報告書. https://www.meti.go.jp/press/2023/08/20230831002/20230831002-1.pdf. 2024 年 11 月 28 日閲覧.