

評価値に翻弄されないレビュー情報探索を促す検索結果スニペット

庵谷 拓輝[†] 山本 祐輔[†]

[†] 静岡大学情報学部 〒432-8011 静岡県浜松市中区城北 3-5-1

E-mail: [†]ihoriya@design.inf.shizuoka.ac.jp, ^{††}yamamoto@inf.shizuoka.ac.jp

あらまし 本稿では、レビューサイトで検索中のユーザに対して、慎重な情報探索を促すスニペットの生成・提示手法を提案する。慎重な情報探索を促すスニペットは、ユーザが損失を感じる情報を加えることによって生成する。生成したスニペットの有効性を検証するために、クラウドソーシングを用いて 200 名の実験協力者を対象にユーザ実験を行った。実験協力者の探索ログ分析から、以下のことが明らかになった：(1) 提案するスニペットを表示するインタフェースは、ユーザの SERP 閲覧時間を増加させる。(2) 提案するスニペットを表示するインタフェースは、ユーザの詳細ページ閲覧時間・閲覧回数、タスク実施時間に影響を与えない。(3) 提案するスニペットを表示するインタフェースは、ユーザの閲覧したアイテムの平均・最低スコアなどの指標に影響を与えない。

キーワード 情報検索, 認知バイアス, ヒューマンファクター

1 はじめに

今日、多くの人々がレビュー情報サイト（以下、レビューサイト）を利用して商品やサービスの購買活動を行っている。レビューサイトには、実際に商品・サービスを購入したユーザによる評価情報が多数投稿されており、評価を示すスコア（以下、評価値）やコメントなどを確認することができる。総務省の調査によると、ウェブで買い物をするユーザの大半がレビューサイトの評価情報を参考にしていることが報告されており、評価情報は購買プロセスに大きな影響を与えている [1]。

一般に、レビューサイトの商品が一覧表示される検索結果ページ（以下、SERP）では、各商品の検索結果概要（以下、スニペット）に評価値が表示される。この評価値は、ユーザの認知バイアスを引き起こし、慎重な購買意思決定を阻害していると考えられる。

実際に、SERP 上のウェブ検索ユーザの行動には、認知バイアスの存在が確認されている。White は、ウェブ検索ユーザは、自身にとって否定的な情報よりも肯定的な情報を好むというバイアスがあることを明らかにした [2]。また、Baeza は、検索結果上位の文書が優先してクリックされるというポジションバイアスの存在を報告している [3]。

レビューサイト上の SERP においても、ユーザは検索順位や評価値が高い商品を優先して閲覧している可能性が高いと考えられる。表層的な情報である検索結果順位や評価値のみで商品の購買意思決定を行うと、自身の重視する観点を見失った短絡的な意思決定に陥る可能性がある。例として、レビューサイトを通じて宿泊施設を予約するケースを考える。SERP 上の評価値のみでは、宿泊施設の評価の理由が「部屋が綺麗であるから」なのか、「温泉が広いから」なのか判断することができない。評価値に囚われた意思決定を行うと、自身の重視する観点に適さない宿泊施設を選ぶ可能性が高くなり、旅行の満足度に大きな影響が生じる。

本稿では、レビューサイトの SERP 上において、評価値による認知バイアスを軽減し、ユーザに慎重な情報探索を促すスニペットの生成手法を提案する。本稿では、評価値による認知バイアスを「ユーザが他ユーザが付与した総合評価スコアのみを重視して情報を取捨選択してしまう傾向」と定義する。SERP 上に表示する各文書のスニペットに、ユーザが損失を感じる情報を加えることで、慎重な情報探索を促すスニペットを生成する。たとえば、宿泊施設レビューサイトにて、あるユーザが評価値を注視してホテルを探しているとする。このユーザが SERP 上で注目している、評価値が 5 であるホテルのスニペットに対して、“夕食”の観点で評判が悪いことを示す情報を提示する。図 1 に、評判の悪い観点情報を追加したスニペットの例を示す。行動経済学の分野では、人間は「利得の喜びよりも、損失の悲しみのほうが大きく感じる」という傾向「損失回避バイアス」を有することが明らかになっている [4]。提案手法によって損失回避バイアスが誘発されることで、ユーザは評価値のみに依存せず、情報を慎重に調べようとするのが期待される。

2 関連研究

2.1 情報検索における認知バイアス

情報検索における認知バイアスについての研究は数多く行われている。White は、ウェブ検索においてユーザが事前に持つ信念が検索行動へ与える影響を調査した [2]。その結果、ユーザは、事前に持つ信念を支持する情報を優先的に閲覧するバイアスを有し、逆の信念を支持する情報を見ようとしなかったことを明らかにしている。Baeza は、ユーザが検索エンジンを利用する際に、SERP 上でランキングが高いもののほどクリック率も高くなる傾向であるポジションバイアスの存在を明らかにしている [3]。このバイアスへの対処のために、Craswell らはポジションバイアスが現実世界にてどのように発生するかを確率モデル化している [5]。また Yue らは、SERP 上の各文書が持つタイトル、URL、サマリーなどの魅力がユーザのクリック行動に与

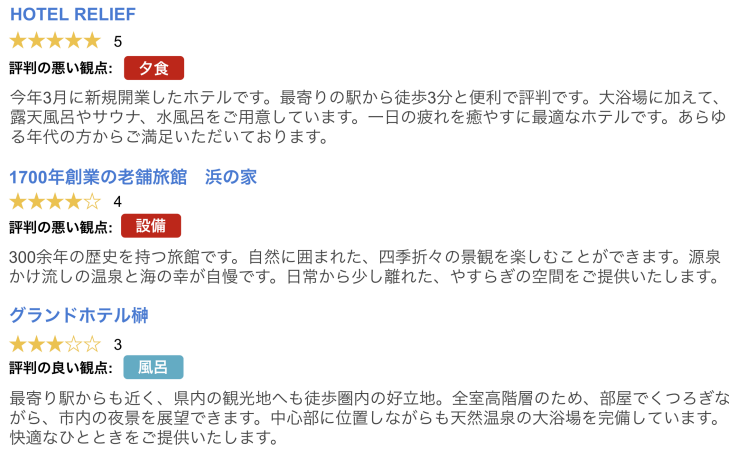


図 1 SERP 上で表示する、損失を感じさせる観点を追加したスニペットの例

える影響を調査した [6]。調査結果から、魅力的なタイトルを持つ検索結果は、ユーザのクリック行動にユーザインタラクション上のバイアスを生じさせることが明らかとなっている。

これらの関連研究をふまえると、ユーザはレビューサイトにおいても、検索順位や評価値が高い商品を優先して閲覧していると考えられる。本稿では、レビューサイト上のポジションバイアスおよび評価値による認知バイアスへの対処のために、情報探索を促す検索結果スニペットを生成・提示手法を提案する。

2.2 商品検索

EC サイトやレビューサイトの普及に伴い、情報検索の一分野として商品検索は注目を集めている [7]。Sondhi らは、EC サイトでの検索ログを分析し、ユーザのクエリと検索行動について調査を行った [8]。調査の結果、ユーザの作成するクエリは 5 つのカテゴリに分類され、それぞれに特徴的な検索行動が見られることを報告している。たとえば、ユーザが比較的高価かつメジャーな商品を探すために特定のクエリを用いて商品検索を行った場合、選択肢は絞られているが、ある程度の調査を必要とする検索を行う傾向にあるとされている。Su らは、EC サイトにて商品検索を行うユーザの検索意図を明らかにするために、ユーザ調査とログ分析を行った [9]。その結果に基づき、ユーザの検索意図を「目標発見」「意思決定」「探索」の 3 つのカテゴリに分類できると報告している。

本稿では、「意思決定」と「探索」に着目したタスクを実験協力者に課し、提案手法の評価を行う。

2.3 情報推薦・検索の結果説明性

情報推薦システムが生成する結果に説明を追加することで、ユーザの推薦結果および推薦システムへの受容性が高くなることが明らかにされている [10]。Tsukuda らは、繰り返し消費されるアイテムである音楽に着目して、個人的要因、社会的要因、アイテム要因に基づいた 9 種類の楽曲推薦の説明手法を提案している [11]。加えて、ユーザ実験によってそれらの説明の有用性の評価を行っている。また Ai らは、商品検索において検索システムと顧客の双方が持つ商品関連性の認識のギャップに着

目した説明可能な検索モデルを提案し、実証実験を行った [12]。

本稿の提案手法では、損失を説明する情報を提示することで、ユーザの慎重な情報探索の促進を狙う。

3 提案手法

3.1 提案システムの概要

本稿では、宿泊施設や家電製品、病院などの商品・サービス（以下アイテム）の検索システムを対象に、慎重な情報探索を促す検索結果スニペット（以下、提案スニペット）を提案する。提案スニペットは、ユーザに検索結果リストに含まれる各アイテムのある観点对する損失を感じさせることで、検索結果順位やレビュースコアといった表層的な情報だけに頼らず、詳細に各アイテムを調べることを促す。本稿では、各アイテムに関する損失を感じさせるために、損失回避バイアスに着目する。損失回避バイアスとは、行動経済学の分野で提唱された、人間は「利得の喜びよりも、損失の悲しみのほうが大きく感じる」という傾向のことである [4]。

本稿では、提案スニペットの適用ドメインとして、宿泊施設レビューサイトの検索システムを設定する。図 1 は、提案システムで、ある地域の宿泊施設を検索した際の出力結果例である。一般的なレビューサイトの検索システムでは、各検索結果に「宿泊施設名」「レビュー総合スコアの平均値」「宿泊施設の概要文」が提示される。これら情報に加えて、提案システムはレビュー総合スコアの下に、損失を感じさせる観点を提示する。提示する観点は、宿泊施設およびそのレビュー総合スコアの多寡によって変化する。レビュー総合スコアが高い場合、宿泊施設の短所を意識させるため、レビューコメントから評判の悪い観点を抽出し提示する。レビュー総合スコアが低い場合、宿泊施設の長所にも目を向けさせるため、レビューコメントから評判の良い観点を抽出し提示する。

例えば、図 1 では、HOTEL RELIEF というホテルの評価値は 5 であり、県下の評価値の平均よりも高い。そのため、「夕食」の評判が悪いことを伝える提案スニペットを提示する。ユーザが提示された評判の悪い観点を見ることによって「夕食の観点

で満足できない可能性がある」という損失を考慮して、該当ホテルの情報をより詳細に調べるようになることが期待される。

3.2 提案スニペットの生成手法

損失回避バイアスを誘発させ慎重な情報検索を促すために、以下の手順で提案スニペットを生成する。

(1) クエリにマッチする宿泊施設のレビューコメントから、エンティティの抽出およびそれに対する感情値の算出を行う

(2) エンティティの出現頻度を考慮して、各宿泊施設の特徴的なエンティティを宿泊施設を評価する「観点」の候補とする

(3) 手順2で抽出した観点候補から、感情値の分布を考慮して、各宿泊施設の短所・長所を考える観点を絞り込む

(4) 宿泊施設のレビュー総合スコアの大小を考慮して、宿泊施設のポジティブな観点、ネガティブな観点のいずれを「損失を感じさせる観点」として提示するかを決定する

(5) 宿泊施設ごとに、損失を感じさせる観点を埋め込んだ検索結果スニペットを生成・提示する

手順1では、レビューコメントに含まれる宿泊施設のある観点に関する評価を抽出するために、Google Natural Language API¹を用いて、宿泊施設のレビューコメントからのエンティティ抽出およびエンティティ感情値の算出を行った。一般に、ユーザによって実際に書き込まれたレビューコメントでは、「このホテルの朝食はとても美味しかった」といったような、ある観点から宿泊施設を評価する言及が行われる。提案システムは、クエリにマッチする各宿泊施設のレビューコメントに対してGoogle Natural Language APIのエンティティ感情分析を適用することで、レビューコメントに表れる各エンティティの感情を表す数値スコア（以下、感情値）を算出できる。ここでエンティティとは、「ホテル」や「朝食」といった名詞を指す。感情値は-1.0から1.0の数値によって表され、-1.0に近いほどネガティブな観点、1.0に近いほどポジティブな観点であるとみなされる。たとえば「朝食がとても美味しかった」というコメントに対して、エンティティ感情分析を行った場合、「朝食」がエンティティとして抽出され、それに対して感情値0.85が得られる。

手順2では、宿泊施設のレビューコメントに出現するエンティティの中から、宿泊施設を評価する上で特徴的なものを「損失を感じさせる観点」の候補として抽出する。エンティティの特徴スコアの計算には、TF-IDF 重み付けを用いる。あるクエリの検索結果リストに含まれる宿泊施設の数 N としたとき、ある宿泊施設 i のレビューコメント集合 $C_i = \{c_{i,1}, c_{i,2}, \dots\}$ に出現するエンティティ e の TF-IDF スコア $tfidf(e, C_i)$ を、以下の式で計算する：

$$tfidf(e, C_i) = tf(e, C_i) \times idf(e) \quad (1)$$

$$tf(e, C_i) = \frac{C_i \text{ における } e \text{ の出現頻度}}{C_i \text{ に関する最大 } tf \text{ 値}} \quad (2)$$

$$idf(e) = \log \frac{N}{df(e)} \quad (3)$$

$$df(e) = \text{レビューコメントに } e \text{ を含む宿泊施設数} \quad (4)$$

各宿泊施設について、レビューコメントに出現するエンティティの $tfidf$ 値を計算し、値の上位10件を「損失を感じさせる観点」の候補として用いる。

手順3では、観点候補をさらに絞り込む操作を行う。提案手法の狙いは、宿泊施設の最も悪い点（あるいは良い点）を提示することで、それを確認しなかったことによる生じうる損失をユーザに感じさせることにある。それゆえ、手順3ではレビューコメントにおける各観点（エンティティ）に対する感情値をふまえて、観点を絞り込む。具体的には、以下の手順で処理を行う。宿泊施設 i の全レビューコメントにおいて出現する観点候補 e の感情値の平均値を計算し、その値を i の感情値とする $(s(e, i))$ 。検索結果に含まれる全宿泊施設 $I = \{i_1, i_2, \dots, i_N\}$ について、手順2に従って観点候補を抽出した後、各宿泊施設の観点候補に対して、感情値を計算する。その後、感情値が正規分布に従うと仮定して、各宿泊施設の各観点の感情値を z スコアに変換する。最終的に、各宿泊施設の最も悪い観点と最も良い観点を抽出するために、各宿泊施設の観点のうち、 z スコアが最大および最小の観点を選択する。

手順4では、宿泊施設のレビュー総合スコアの大小を考慮して、各宿泊施設に対してポジティブな観点とネガティブな観点のいずれを表示するかを決定する。あるクエリに対する検索結果リストに含まれる宿泊施設のレビュー総合スコアの平均値を r_{avg} とする。本稿では、

- ある宿泊施設 i のレビュー総合スコアが r_{avg} より大きいときは、 e のネガティブな観点
- ある宿泊施設 i のレビュー総合スコアが r_{avg} より小さいときは、 e のポジティブな観点を提示する。

手順5では、図1のように手順4で求めた宿泊施設語の損失を感じさせる観点を含めたスニペットを生成・提示する。

3.3 仮 説

提案スニペットを提示することで、ユーザに損失回避バイアスが生じることが期待される。これによって、ユーザは評価値という表層的な情報に翻弄されず、より慎重に宿泊施設を探索すると考えられる。本稿では、以下の仮説を検証する。

H1 SERP 上に提案スニペットを提示することで、ユーザの詳細ページ閲覧時間・閲覧回数とタスク所要時間は増加する

H2 SERP 上に提案スニペットを提示することで、ユーザは、より積極的に評価値以外の宿泊施設に関する情報を調べるようになる

1 : <https://cloud.google.com/natural-language/>

4 実 験

4.1 実験協力者

クラウドソーシングサイトの Lancers²を利用して、全 200 名の実験協力者を募った。本実験は 2021 年 1 月に実施され、タスクを適切に完了した実験協力者に対して、150 円の報酬を支払った。本実験では、各実験協力者に 4 つの探索タスクを課した。全 200 名の実験協力者から得られたデータから、タスクを途中で終了した実験協力者やログが正しく計測されていない実験協力者などのデータを除去した。最終的に、178 名の実験協力者のデータを分析に利用している。

4.2 検索結果インタフェース

本実験では、以下の 4 種類の検索結果インタフェースを用意した。3.2 節と同様に、あるクエリに対する検索結果リストに含まれる宿泊施設のレビュー総合スコアの平均値を r_{avg} とした。**CorrUI** “ r_{avg} に対する、ある宿泊施設のレビュー総合スコアの大小”と“表示する観点の感情の度合い”を相関させるインタフェース

InvCorrUI “ r_{avg} に対する、ある宿泊施設のレビュー総合スコアの大小”と“表示する観点の感情の度合い”を逆相関させるインタフェース

BothUI ポジティブ、ネガティブ双方の観点を表示するインタフェース

NoneUI 観点を表示しないインタフェース

1 つ目は、ある宿泊施設のレビュー総合スコアが r_{avg} より大きい場合はポジティブな観点を表示し、小さい場合はネガティブな観点を表示するインタフェースである。本稿では、これを **CorrUI** と呼ぶ。例えば、ある宿泊施設のレビュー総合スコアを 4.0、 r_{avg} を 3.5 とする。ある宿泊施設のレビュー総合スコアは r_{avg} よりも高いため、**CorrUI** ではポジティブな観点を表示する。

2 つ目は、ある宿泊施設のレビュー総合スコアが r_{avg} より大きい場合はネガティブな観点を表示し、小さい場合はポジティブな観点を表示するインタフェースである。本稿では、これを **InvCorrUI** と呼ぶ。これが本稿で提案する、損失回避バイアスの誘発を狙った検索結果スニペットを表示するインタフェースである。上述の例と同様に、ある宿泊施設のレビュー総合スコアを 4.0、 r_{avg} を 3.5 とする。ある宿泊施設のレビュー総合スコアは r_{avg} よりも高いため、**InvCorrUI** ではネガティブな観点を表示する。

3 つ目は、各ホテルに紐づくポジティブな観点とネガティブな観点の両方を提示するインタフェースである。宿泊施設のレビュー総合スコアと r_{avg} の大小関係を考慮せず、ポジティブおよびネガティブである観点の両方を表示するインタフェースとして、**BothUI** と呼ぶ。

4 つ目は、観点を表示しないベースラインインタフェースである。提案する観点の表示を行わないインタフェースとして、

NoneUI と呼ぶ。このインタフェースは、観点の有無の比較に用いる。提案する観点を含まない一般的な検索結果スニペットを表示する。

観点を表示するインタフェースにおいて、図 1 のように評判の良い（ポジティブな）観点を水色、評判の悪い（ネガティブな）観点を赤色で色付けすることによって目立つようにする。実験協力者は、4 つの探索タスクを通して、上述した全 4 種類の検索結果インタフェースを一度ずつ利用した。

4.3 探索タスク

本実験では、宿泊施設の決定を目的とした探索タスクを合計 4 つ用意した。宿泊施設を題材として扱う理由は、生活の要素となる衣食住に関わり、快適な環境を選ぶための探索行動が必要と考えられるからである。

各タスクに対して、宿泊施設を探す対象の都道府県を設定した。実験協力者の関心がある可能性が高い都道府県を用いるため、「じゃらん宿泊旅行調査 2020」の都道府県別の延べ宿泊旅行者数（推計値）[13] を参考とする。この調査結果の上位 4 都道府県である「東京都・北海道・大阪府・静岡県」を対象の都道府県として選択した。

実験協力者は実験全体を通して、宿泊施設を決定する探索タスクに合計 4 回取り組んだ。

4.4 実験システム

実験協力者が探索タスクを行うためのシステム（以下、実験システム）を構築した。実験協力者が主に利用する画面は、SERP および SERP からリンクされた宿泊施設の詳細情報が閲覧できるページ（以下、詳細ページ）である。

SERP には、ある都道府県の宿泊施設を検索するクエリに基づいた検索結果リストを表示する。各探索タスクにおいて、宿泊施設を探すためのクエリをすでに発行したことを前提とする。検索結果リストに表示するデータには、じゃらん³の宿泊施設データセットを利用している。各探索タスクにおいて、検索結果リストとして最大 50 件のアイテムを用意し、1 ページあたりに 10 件のアイテムを表示する。表示するアイテムの順序は、レビュー総合スコアの高い順としている。各アイテムの要素として、宿泊施設名、レビュー総合スコア、宿泊施設の説明文、適用中のインタフェースに基づく観点を表示する。本稿では、探索タスク中に外部の検索エンジンを利用することを許可していない。

詳細ページには、下記の情報を表示する。

- レビュースコア（総合・食事や部屋などの観点別）
- 基本情報（名称、価格、説明文、種類、県、エリア、住所）
- じゃらんのサイトを表示するインラインフレーム要素
- カスタマーレビュー

図 2 が実際の詳細ページである。宿泊施設の情報を補完するために、インラインフレーム要素（iframe）を用いて実際のじゃらんのページを表示する。実験協力者が探索タスク中に離脱しないように、インラインフレーム要素から外部へのリンク

2 : <https://lancers.jp/>

3 : <https://www.jalan.net/>



図 2 実験システム内の詳細ページ画面

ができないようにしている。実験協力者は、この詳細ページにて宿泊施設を決定し、各探索タスクを完了した。

4.5 実験手順

本実験では、前述の 4 種類の検索結果インタフェースを被験者内要因とする。グレコラテン方格を用いてタスク順序の無作為化を行うことによって、利用するインタフェースやタスクの順序による影響を可能な限り取り除いた。実験は、下記の構成で行った。

- (1) 実験の概要説明
- (2) 探索タスク
- (3) 事後アンケート

まず、実験協力者に実験の概要について説明した。探索タスク中に収集するデータの利用についての同意を得られる者のみ、実験に進むよう指示した。

次に、実験システムを用いて全 4 回の探索タスクに取り組んでもらう。各探索タスクを開始する前に、具体的なタスクの内容を説明するページを設けた。このページにて、実験協力者に下記の指示文を提示した。

連休が取れたため、以前から行きたいと思っていた札幌市への旅行を計画しています。次のページから、旅行に利用する宿泊施設を一件選んでください。

実験協力者は、上記の指示文を読んだ後に探索タスクを開始する。各探索タスクは、詳細ページ上で宿泊施設を決定することによって完了する。実験協力者に探索タスクに真剣に取り組んでもらうため、宿泊施設を決定する際に、その理由を 30 文字以上で入力してもらった。

全 4 つの探索タスクの終了後、事後アンケートに回答してもらった。事後アンケートでは、人口統計的属性に加えて、タスク中の探索態度および日常での評価値の参考度合いを質問した。具体的には、以下の質問に回答してもらった。

- Q1** タスクを通して、検索結果画面に表示された観点（例：「評判の悪い観点：設備」）を意識したか？
- Q2** タスクを通して、宿泊施設を選ぶ際にどのような点を重視したか？
- Q3** 普段インターネットで宿泊施設やレストランを探したり、

買い物をしたりする際に「星などで表示されるレビュースコア」をどの程度参考にするか？

Q1 は選択式、Q2 は自由記述式で回答してもらった。Q3 は 5 段階のリッカート尺度を用いて回答してもらった。行動指標から得られるデータの補完を目的に、上記の質問を行った。

4.6 測定した行動指標

仮説に対応した評価を行うため、探索タスク中の実験協力者から以下のデータを収集した。

SERP 閲覧時間 タスク中の検索結果ページの合計閲覧時間

詳細ページ閲覧時間 タスク中の詳細ページの合計閲覧時間

詳細ページ閲覧回数 タスク中に詳細ページの合計閲覧回数

最大クリック深度 SERP において、最も低い位置でクリックされたページの順位

タスクの所要時間 タスクの開始から終了までにかかった時間

閲覧された宿泊施設の平均スコア タスク中に閲覧された宿泊施設のレビュー総合スコアの平均

閲覧された宿泊施設の最低スコア タスク中に閲覧された宿泊施設の中で、最も低いレビュー総合スコア

決定された宿泊施設のスコア タスク終了時に決定された宿泊施設のレビュー総合スコア

仮説 H1 および H2 の立証のために、上記の行動指標を用いて分析を行った。

5 結果

本章では、実験協力者計 178 名のタスク中の行動ログの分析結果を述べる。各行動指標にて、正しく計測されていないデータは分析対象から除外している。実験にて設定した 4 つの探索タスクは、すべて宿泊施設を決定するというものであり、タスク間に差はないものと仮定している。要因をインタフェースとして、その水準を 4.5 節で述べた四種のインタフェースとしている。ユーザ実験で収集したデータには、分布の正規性が確認されなかった。そのため、統計的有意差の検証にはノンパラメトリックな検定法を利用している (有意水準 $\alpha = 0.05$)。

5.1 SERP 閲覧時間

実験協力者がどの程度の時間を使って、SERP(検索結果一覧ページ) 上で次に閲覧するページを選んだか調べるため、SERP 閲覧時間 (秒) を分析した。

表 1 に、インタフェースごとの SERP 閲覧時間の平均値と標準偏差を示す。表 1 によると、SERP 閲覧時間において平均値が最も高いのは InvCorrUI の 102.29 であり、最も低いのは CorrUI の 78.24 である。

インタフェース間に差があるかを調べるために、Friedman 検定を行った。その結果、表 1 に示すように統計的有意差が確認された ($p = 0.007$)。加えて、どのインタフェース間に差があるかを調べるために、Wilcoxon の符号付き順位検定を行った。この際、p 値の補正のために Bonferroni 法を利用した。その結果、表 2 に示すように、InvCorrUI と NoneUI 間に統計的有意差があることが確認された ($p = 0.0024$)。

表 1 行動指標の分析結果 (上段: 平均値, 下段: 標準偏差)

行動指標	UI 条件				<i>p</i> -value
	CorrUI	InvCorrUI	BothUI	NoneUI	
SERP 閲覧時間 (秒)	78.24 (107.72)	102.29 (135.09)	100.97 (217.82)	99.88 (266.41)	0.0007
詳細ページ閲覧時間 (秒)	213.87 (197.97)	219.01 (215.47)	207.13 (173.35)	226.92 (226.15)	0.36
詳細ページ閲覧回数	5.08 (6.70)	4.70 (5.74)	4.48 (5.90)	4.96 (5.87)	0.12
最大クリック深度	13.86 (13.63)	14.20 (14.13)	12.37 (11.85)	15.21 (14.63)	0.09
タスク所要時間 (秒)	335.58 (276.16)	379.65 (360.52)	381.87 (578.97)	386.20 (422.68)	0.41
閲覧した宿泊施設の平均スコア	4.41 (0.29)	4.42 (0.30)	4.44 (0.31)	4.40 (0.32)	0.3
閲覧した宿泊施設の最低スコア	4.23 (0.48)	4.24 (0.41)	4.29 (0.47)	4.22 (0.46)	0.07
決定した宿泊施設のスコア	4.40 (0.317)	4.41 (0.32)	4.41 (0.36)	4.38 (0.33)	0.52

CorrUI では、スコアの高い宿泊施設に対して評判の良い観点を表示した。すなわち、実験協力者が CorrUI を用いた場合には、他のインタフェースと比較して、SERP の閲覧にあまり時間をかけなかったことが分かる。また、実験協力者が InvCorrUI を用いた場合、特に NoneUI を用いた場合と比較して、時間をかけて SERP を眺めていたことがうかがえる。このことから、提案スニペットを表示する InvCorrUI を用いた場合、探索行動中の SERP 閲覧時間が長くなることが明らかとなった。この結果は、仮説 H2 を支持する。

5.2 詳細ページ閲覧時間

実験協力者が各宿泊施設を決定するために、どの程度時間を使っていたか調べるため、詳細ページ閲覧時間 (秒) を分析した。

表 1 に、インタフェースごとの詳細ページ閲覧時間の平均値と標準偏差を示す。表 1 によると、インタフェース間で最も大きな平均値は NoneUI の 226.92 である。

インタフェース間に差があるかを調べるため、Friedman 検定を行った。その結果、表 1 に示すように、統計的に有意な差は確認されなかった ($p = 0.36$)。

このことから、インタフェースの違いによって、詳細ページ閲覧時間に差は生じないことが明らかとなった。すなわち、仮説 H1 は支持されない。

5.3 詳細ページ閲覧回数

実験協力者がどの程度参考情報となる詳細ページを閲覧したか調べるため、詳細ページ閲覧回数を分析した。

表 1 に、インタフェースごとの詳細ページ閲覧回数の平均値と標準偏差を示す。表 1 によると、最も大きな平均値は CorrUI の 5.08 であり、最も小さい平均値は InvCorrUI の 4.7 である。

インタフェース間に差があるかを調べるため、Friedman 検定を行った。その結果、表 1 に示すように、統計的に有意な差は確認されなかった ($p = 0.12$)。

このことから、インタフェースの違いによって、詳細ページ閲覧回数に差は生じないことが明らかとなった。この結果は、InvCorrUI によって詳細ページ閲覧回数が増加しないことを意味する。すなわち、仮説 H1 は支持されない。

5.4 最大クリック深度

実験協力者がどの程度 SERP を走査したか分析するために、最も低い位置でクリックされたページの順位として最大クリック深度を分析した。

表 1 に、インタフェースごとの最大クリック深度の平均値と標準偏差を示す。表 1 によると、最も大きな平均値は NoneUI の 15.21 であり、最も小さい平均値は BothUI の 12.37 であることが分かる。

インタフェース間に差があるかを調べるため、Friedman 検定を行った。その結果、表 1 に示すように、有意水準 10% で統計的有意傾向が確認された ($p = 0.09$)。

しかし、表 1 に示すように、最大クリック深度の大きいものは NoneUI である。仮説 H2 では、InvCorrUI を用いた場合に最大クリック深度が大きくなることを期待していた。すなわち、仮説 H2 は支持されない結果となった。

5.5 タスク所要時間

実験協力者がタスクの完了にどの程度時間を費やしたかを調べるために、タスク所要時間 (秒) を分析した。

表 1 に、インタフェースごとのタスク所要時間の平均値と標準偏差を示す。表 1 によると、最も大きな平均値は NoneUI の 386.20 であり、最も小さい平均値は CorrUI の 335.68 である。

インタフェース間に差があるかを調べるため、Friedman 検定を行った。その結果、表 1 に示すように、統計的に有意な差は確認されなかった ($p = 0.41$)。

このことから、インタフェースの違いによって、タスク所要時間に差は生じないことが明らかとなった。この結果は、

InvCorrUI によってタスク所要時間が増加しないことを意味する。すなわち、仮説 H1 は支持されない。

5.6 閲覧した宿泊施設の平均スコア

インタフェースが実験協力者の閲覧した宿泊施設のスコアに影響を与えていたかどうか調べるため、閲覧した宿泊施設の平均スコアを分析した。

表 1 に、インタフェースごとの閲覧した宿泊施設の平均スコアの平均値と標準偏差を示す。表 1 によると、インタフェース間で、最も大きい平均値は BothUI の 4.44 であり、最も小さい平均値は NoneUI の 4.40 である。

インタフェース間に差があるかを調べるため、Friedman 検定を行った。表 1 に示すように、統計的に有意な差は確認されなかった ($p = 0.3$)。

このことから、インタフェースの違いによって、閲覧した宿泊施設の平均スコアに差は生じないことが明らかとなった。すなわち、仮説 H2 は支持されない。

5.7 閲覧した宿泊施設の最低スコア

インタフェースが実験協力者の閲覧した宿泊施設のスコアに影響を与えていたかどうか調べるため、閲覧した宿泊施設の最低スコアについても分析した。

表 1 に、インタフェースごとの閲覧した宿泊施設の最低スコアの平均値と標準偏差を示す。表 1 によると、インタフェース間で最も大きい平均値は BothUI の 4.29 であり、最も低い平均値は NoneUI の 4.22 である。

インタフェース間に差があるかを調べるため、Friedman 検定を行った。表 1 に示すように、有意水準 10% で統計的に有意傾向が確認された ($p = 0.07$)。

しかし、表 1 に示すように、閲覧した宿泊施設の最低スコアが小さいものは NoneUI である。仮説 H2 では InvCorrUI を用いた場合に、閲覧した宿泊施設の最低スコアが小さくなることを期待していた。すなわち、仮説 H2 は支持されない結果となった。

5.8 決定した宿泊施設のスコア

インタフェースが実験協力者の最終的に選んだ宿泊施設のスコアに影響を与えていたかどうか調べるため、決定した宿泊施設のスコアを分析した。

表 1 に、インタフェースごとの決定した宿泊施設のスコアの平均値と標準偏差を示す。表 1 によると、インタフェース間で、最も大きい平均値は BothUI の 4.41 であり、最も小さい平均値が NoneUI の 4.38 であることが分かる。

インタフェース間に差があるかを調べるため、Friedman 検定を行った。表 1 に示すように、統計的に有意な差は確認されなかった ($p = 0.52$)。

このことから、インタフェースの違いによって、決定した宿泊施設のスコアに差は生じないことが明らかとなった。すなわち、仮説 H2 は支持されない。

表 2 SERP 閲覧時間における多重比較の結果 (p -value, Bonferroni の補正後の有意水準: 0.0083)

	BothUI	CorrUI	InvCorrUI
CorrUI	1.0000	-	-
InvCorrUI	1.0000	0.0094	-
NoneUI	0.4964	1.0000	0.0024

6 考察

6.1 結果の解釈と仮説の立証結果

提案スニペットを表示する InvCorrUI の利用は、ユーザの探索行動中の SERP 閲覧時間を長くすることが明らかになった。この理由は、ユーザが SERP 上に表示された観点に注目し、次にどのページを開くか考える時間が伸びたからである考えられる。

一方、詳細ページ閲覧時間・閲覧回数をはじめとした、その他の指標では仮説 H1 および H2 を支持する有意な差は確認されなかった。

仮説 H1 として、提案スニペットを表示する InvCorrUI を利用することで、ユーザの詳細ページ閲覧時間・閲覧回数とタスク所要時間は増加すると予想した。詳細ページの閲覧時間・閲覧回数およびタスク所要時間にて、インタフェース間で有意な差が得られず、仮説 H1 は支持されなかった。

仮説 H2 として、提案スニペットを表示する InvCorrUI を利用することで、ユーザはより積極的に評価値以外の宿泊施設に関する情報を調べるようになることが予想した。SERP 閲覧時間の指標では、提案スニペットを表示する InvCorrUI を利用することで、閲覧時間が長くなるという効果の発揮が確認された。この結果より、仮説 H2 は部分的に支持されたといえる。

6.2 提案手法の限界

6.2.1 観点抽出の精度

提案手法が詳細ページ閲覧回数や最大クリック深度に影響を与えられなかったことから、検索結果スニペットに付与する観点の抽出精度が課題として挙げられる。その理由は、SERP に表示された観点が正確に宿泊施設の特徴を表している場合、ユーザがより検索結果が下位であるアイテムをクリックする確率が高くなる可能性があるからである。

観点抽出の精度を高めるに、より適切なアルゴリズムの試行と選択を行う必要がある。

6.2.2 適用ドメインの違いによる意思決定の性質差

ユーザ実験での提案手法の適用ドメインが、宿泊施設のみに限定されていることが課題として挙げられる。レビューサイトを利用した意思決定は、ドメインが異なることで、リスク度合いに差が生じる。例えば、レビューサイトにて病院を探す場合と宿泊施設を探す場合では、失敗した場合のリスク度合いが異なる。病院の選択は健康状態に関わるため、宿泊施設のそれと比較して、重大な意思決定となる。本稿のユーザ実験における探索対象は宿泊施設のみであるため、より重大な状況での意思

決定を考慮できていない。適用ドメインの違いによって、提案手法の効果の発揮度合いに差が現れるか、調査が必要である。

6.2.3 アンケート設計の改良

実験協力者には、ユーザ実験の後に 4.5 節で説明した事後アンケートを行ってもらった。この事後アンケートにて、用意した各インタフェースに対する適切なフィードバックが得られていないことが課題として挙げられる。ユーザ実験において、実験協力者は合計 4 つのタスクにて、異なるインタフェースを利用している。したがって、それぞれのインタフェースに対するフィードバックを得るために、各タスクの終了後にアンケートを設けるべきであった。また、それぞれのインタフェースを利用した際に、宿泊施設を決定する上で、評判の良い点のみならず悪い点についても考慮していたかどうか調査すべきであった。これらについて調べることで、ユーザの探索行動中の意識をより詳細に知ることができる。

ユーザの探索行動中の意識を詳細に調査するために、アンケート設計を改良して再実験を行う必要がある。

7 結 論

本稿では、レビューサイトで検索中のユーザに対して、慎重な情報探索を促すスニペットの生成・提示手法を提案した。SERP 上で表示される検索結果スニペットに、損失回避バイアスを誘発する観点情報を付与することで、ユーザの慎重な情報探索の促進を狙った。

提案手法の有効性検証のために、200 名の実験協力者を対象に、宿泊施設を題材とした探索タスクを課すユーザ実験を実施した。実験協力者の探索ログを用いて、設定した測定指標にて分析を行った。その結果、提案スニペットを表示するインタフェースの利用は、ユーザの SERP 閲覧時間を伸ばす効果があることが判明した。一方で、詳細ページ閲覧時間・閲覧回数などのその他の指標に、影響を与えないことが分かった。

今後は、より正確な観点情報を SERP 上で提示するために、提案スニペットの生成過程を改善し、観点情報の抽出精度を高める必要がある。また、探索対象ドメインの追加とアンケート設計の改良を行い、ユーザ実験を再実施する必要がある。

謝 辞

本研究は JSPS 科研費 JP18KT0097, JP18H03243, JP18H03494, C18H032440 および課題設定による先導的人文・社会科学研究推進事業の助成を受けたものです。ここに記して謝意を表します。

文 献

- [1] 総務省. 平成 28 年版 情報通信白書, 2016. <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h28/index.html>, (2020 年 12 月 14 日閲覧).
- [2] Ryen White. Beliefs and biases in web search. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pp. 3–12, 2013.

- [3] Ricardo Baeza-Yates. Bias on the web. *Communications of the ACM*, Vol. 61, No. 6, pp. 54–61, 2018.
- [4] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*, pp. 99–127. World Scientific, 2013.
- [5] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. An experimental comparison of click position-bias models. In *Proceedings of the 2008 international conference on web search and data mining*, pp. 87–94, 2008.
- [6] Yisong Yue, Rajan Patel, and Hein Roehrig. Beyond position bias: Examining result attractiveness as a source of presentation bias in clickthrough data. In *Proceedings of the 19th international conference on World wide web*, pp. 1011–1018, 2010.
- [7] Ilya Markov and Maarten de Rijke. What should we teach in information retrieval? In *ACM SIGIR Forum*, Vol. 52, pp. 19–39. ACM New York, NY, USA, 2019.
- [8] Parikshit Sondhi, Mohit Sharma, Pranam Kolari, and ChengXiang Zhai. A taxonomy of queries for e-commerce search. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 1245–1248, 2018.
- [9] Ning Su, Jiyin He, Yiqun Liu, Min Zhang, and Shaoping Ma. User intent, behaviour, and perceived satisfaction in product search. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 547–555, 2018.
- [10] Jonathan L Herlocker, Joseph A Konstan, and John Riedl. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pp. 241–250, 2000.
- [11] Kosetsu Tsukuda and Masataka Goto. Explainable recommendation for repeat consumption. In *Fourteenth ACM Conference on Recommender Systems*, pp. 462–467, 2020.
- [12] Qingyao Ai, Yongfeng Zhang, Keping Bi, and W Bruce Croft. Explainable product search with a dynamic relation embedding model. *ACM Transactions on Information Systems (TOIS)*, Vol. 38, No. 1, pp. 1–29, 2019.
- [13] リクルートじゃらんリサーチセンター. じゃらん宿泊旅行調査 2020, 2020. https://jrc.jalan.net/surveys/accommodation_travel/, (2021 年 1 月 14 日閲覧).