オンラインショッピングにおける満足度と検索行動の関係分析

柳田 雄輝[†] 加藤 誠^{††} 河田 友香^{†††} 山本 岳洋^{†††} 大島 裕明^{†††} 藤田 澄男^{††††}

† 筑波大学大学院 情報学学位プログラム〒 305-8550 茨城県つくば市春日 1-2†† 筑波大学 図書館情報メディア系〒 305-8550 茨城県つくば市春日 1-2††† 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†††† ヤフー株式会社〒 102-8282 東京都千代田区紀尾井町 1-3

E-mail: †ynagi2@klis.tsukuba.ac.jp, ††mpkato@acm.org,

あらまし 本研究では Web 情報探索行動と購買満足度の関係を調査する. 我々は Web 情報探索行動の記録として Web 検索ログを, 購買満足度の評価として EC サイトにおける商品レビューの点数を用いて分析を行った. 分析の結果, 満足したユーザは不満足なユーザに比べ, (1) 購買前1週間以内に, より幅広く商品に関連する情報を検索する頻度が高いことがわかった. (2) 特に Web 検索に比較的慣れている場合や比較的高価な商品を購買する場合において, 購買前に具体的なクエリで多く検索していた. さらに, 購買前後の Web 検索行動から購買後の満足度を予測する問題にも取り組み、顧客の Web 検索の種類が異なれば、購買満足度も異なる可能性があることが示された.

キーワード ユーザモデル・行動分析,情報探索行動,検索クエリ,検索意図,購買満足度

1 はじめに

近年オンラインショッピングにて商品を購買する人々は増加しており、その利用者を満足させることは重要な課題として認知されている。PwCの調査によると、週に1回以上スマートフォンなどのモバイル端末から商品を購買している人の割合は、2019年は24%であったが、2021年12月には41%とその割合が増加している[1]。店頭でのショッピングに対し、オンラインショッピングの体験はWebでの情報行動の影響を強く受けると考えられる。そのため、ECサイトではこのようなオンラインショッピングの利用者を満足させる、より良いデジタル体験を設計することが重要な課題の1つとなっている。

このような背景から購買満足度に関する研究は広く行われているが、商品購買者の Web 検索による情報探索行動と購買満足度の関係についてはまだ十分な理解が得られていない. 既存の購買満足度に関する研究では、Web 上での情報探索行動が考慮されていない. 例えば、商品に対する関心の高さ [2] や知覚された商品の質 [3] は購買満足度に影響することが知られている. その一方で、Web 上での情報探索行動が購買満足度に及ぼす影響については、まだ広く調査されていない. 最も関連性の高い研究として、Su らによる、EC サイトでの検索行動と検索結果に対する満足度の関係についての研究が挙げられる [4]. しかし、この研究では 2 つの重要な疑問が解決されていない. 1 つは、Web 検索行動と顧客満足度の関係である. Web サーチエンジンは主要な情報収集手段であり、その効果はオンラインショッピングにおいても無視できない. もう 1 つは、検索行

動と購買満足度の関係である。検索結果の満足度が必ずしも商品購買の満足度を示すとは限らないので、商品購買後の満足度に対する効果を明らかにするためには、さらなる調査が必要である。

本研究では、Web 上での情報探索行動と購買満足度の関係を分析する。我々はこれまでに、カメラカテゴリを対象としてWeb 検索による情報探索行動と購買満足度の関係を分析してきた[5],[6]. 本論文では、これらの参考論文で取り組んだ内容を踏まえ、カメラカテゴリ以外のカテゴリも含めた分析を行う他、クエリ語と満足度の関係についても調査する。オンラインショッピングにおける情報探索行動の記録としてWeb 検索ログを用い、購買満足度を表す評価としてECサイトにおける商品レビューの点数を用いて分析を行った。本研究のリサーチクエスチョンは次の通りである:

RQ1. 満足したユーザと不満足なユーザで Web 検索行動は異なるのか?

RQ2. 商品の特性が変われば、満足したユーザと不満足なユーザで Web 検索行動は異なるのか?

RQ3. ユーザの特性が変われば、満足したユーザと不満足な ユーザで Web 検索行動は異なるのか?

RQ4. 購買前, または, 購買後の検索行動から購買満足度は予測できるか?

データセットには日本の大手商用サーチエンジンである Yahoo!検索の Web 検索ログと日本の大手 EC サイトである Yahoo!ショッピングの購買ログおよびレビューデータを用いた. RQ1 に対しては、EC サイトで商品を購買したユーザを特定し、 その検索意図に基づいて Web 検索クエリを分類、購買満足度 が異なるユーザの時間分布を調査した. RQ2, RQ3 の回答として, RQ1 で用いたデータを商品特性, ユーザ特性に着目して掘り下げ, Web 検索の違いを明らかにした. また, 顧客が検索に用いるクエリ語についても分析した. 最後に, RQ4 への回答として, Web 検索の違いが購買満足度を説明できるかどうか調査した. この目的のため, これまでの分析で見出された潜在的な影響力を持つ要因に基づいて, 2 値分類器を学習させた.

本研究の主な貢献は次の通りである:

- まだ研究されていない EC サイト利用者の Web 情報探索行動と購買満足度の関係を調査した.
- SAT (満足した) 顧客と DSAT (不満足な) 顧客, さらには顧客のタイプや商品によって, Web 検索行動が異なることを明らかにした. 特に, (1) 購買前 1 週間以内に, SAT ユーザは DSAT ユーザに比べ, より幅広く商品に関連する情報を検索する頻度が高いことがわかった. (2) SAT ユーザは DSAT ユーザよりも購買前に具体的なクエリで多く検索しており, 特に Web 検索に比較的慣れている場合や比較的高価な商品を購買する場合, DSAT ユーザよりも多く TF クエリで検索している. (3) 他人の意見を求めている顧客は, 購買時に満足する可能性が高い.
- また, 購買前後の Web 検索行動から, 顧客の購買満足度を予測する問題にも取り組んだ. この試みにより, 顧客が行った Web 検索の種類が異なれば, 購買満足度も異なる可能性があることが実証された.

本論文の構成は次の通りである. 2節では購買満足度やユーザの検索行動に関する関連研究について述べる. 3節では,分析に用いたデータの詳細と,Webクエリに基づくユーザの検索意図の推定方法について説明する. 4節では,満足度の異なるユーザのWeb検索行動の違いについて分析した結果を報告する. 5節では,4節で得られた結果をもとに,購買満足度の予測に取り組む. 6節ではインプリケーションと結論を述べる.

2 関連研究

本節では顧客満足度とユーザ検索行動の 2 トピックそれぞれの関連研究について述べる.

2.1 顧客満足度

Razak らは既存の顧客満足度に関する文献について触れた上で,顧客満足度を"ある商品やサービスを購買したあとに顧客の期待と比較され,評価としてフィードバックされるもの"とした[7]. したがって,顧客に知覚された商品やサービスの効用が顧客の期待ほどでなければ顧客は不満を持ち,効用が期待以上であれば顧客は満足する。そして,顧客満足度は売り手の収益やサービスの利用継続などに関係することが示された[8],[9]. 例えば Hallowell は,顧客満足度はロイヤリティに結びつき,ロイヤリティは収益に関係することを明らかにした[8].

満足度に影響を及ぼす要因についても研究されてきた [10], [11], [12]. Cardozo は、顧客満足度は商品そのものだけでなく、商品の獲得に関わる顧客の経験にも依存するとし

た[13]. Tsiotsou は知覚された商品の品質と満足度が購買に与える影響を調査し、知覚された商品の質が顧客満足度や商品の再購買に影響を及ぼすことを明らかにした[3]. Wang らは ECサイトのレビューを用いて、顧客満足度に影響を及ぼす商品特性を分析し、高価格帯の商品は低価格帯の商品と比べ、多くの商品特性が満足度に影響することを明らかにした[14]. Richinsと Bloch は耐久消費財購買後の顧客満足度とその時間的な変化について分析し、商品に対する関心の高い消費者の方が満足度がわずかに高いことを明らかにした[2].

マーケティング研究の文献において、購買前の情報探索は内部探索と外部探索に分けられる. 内部探索とは、顧客自身が以前獲得した情報について、記憶の中から探すことを指す [15]. 対して外部探索とは、広告や知人など、顧客自身の外部から情報を取得して探すことを指す [16]. Punj と Staelin は外部探索量、コスト削減、顧客満足度の関係性を調査し、探索量はコスト削減に有効であるが、満足度はコスト削減にのみ関係することが示された [17]. したがって、この研究では外部探索量と顧客満足度の間には間接的な関係しか見出せなかった. さらに、外部探索量や知覚リスクに影響を及ぼす要因を明らかにする研究が広く行われてきた [18], [19]. 本研究は外部探索が顧客満足度に与える影響に着目した研究に分類される.

このように、商品の質や価格、商品に対する関心の高さや外部探索量が購買満足度に影響を及ぼすことが明らかにされているが、Web 検索による情報探索行動と結びつけた分析はされていない。Punj と Staelin [17] の研究は本研究と最も関連する研究の1つであると考えられるが、彼らの研究での情報探索行動は主に人と話したり、本や雑誌を読んだりすることを指す。このような情報探索行動には、情報ニーズの明示的な表現(i.e.,クエリ)が必ずしも含まれておらず、オフラインで行われることがほとんどである。本研究では、オンラインでの情報探索行動を分析することで、Web 検索に特化した要因とオンラインショッピングに特化した要因の両方を見出すとともに、オンラインショッピングに関する既存研究を支持する知見の両方が得られると期待される。

2.2 ユーザ検索行動

ユーザ検索行動,特に購買に関する検索行動の分析を行った研究もなされている [20], [21], [22], [23], [24].中でも Su らは, Target Finding (TF), Decision Making (DM), Exploration (EP) の 3 種類の検索意図を考案し, EC サイトにおける検索行動を調査するために,検索セッションから検索意図を推定した [4].彼らの分析では,ユーザの検索行動とその背景にある検索意図,検索結果に対する満足度の関係に着目している.Su らの分析は本研究の分析と似ているが,前述したように大きな違いが 2 点ある.Su らの分析における満足度とは,ユーザスタディによって得られた検索体験の満足度のことである.また,検索満足度と購買行動の関係については,"(検索)満足度が必ずしも購買に結びつくとは限らない"と主張しており,購買は検索満足度のシグナルとはなりえないとしている.対して本研究では、ユーザが商品に付与したレーティングを用いることで、

表 1 データの統計情報

X I / / WITH INTE					
	検索ログ	購買ログ	レビュー		
サービス	Yahoo!検索	Yahoo	!ショッピング		
期間	2016 -	2017	2016 - 2018		
ユーザ数	13,882				
購買件数	N/A	満足: 12,6	20; 不満足: 2,657		
クエリ数	48,758,880	N/A			
平均クエリ長	1.70	N/A			
平均セッション長	3.03		N/A		

ユーザの商品購買後も含めた検索行動と商品購買後に付与される購買満足度の関係を分析する. Zhang らはサーチエンジンの広告に対するユーザの広告消費行動を調査し, 購買した商品が広告に表示されている場合, 購買前と購買後で行動に違いがあることを明らかにした [25]. Zhang らは広告消費行動の分析をしていることに対し, 本研究では購買満足度の分析を行う点で異なる.

検索行動の理解にあたり、検索意図の予測や分類法を提案する研究もされている [26], [27]. Web 検索意図の分類として、Broder によって提案された代表的な Web 検索意図の分類法は Navigational, Informational, Transactional からなる [28]. Jansen らは Broder の分類法に基づいて Web クエリを分類した [29]. Rao らは Broder の分類に加え、Web 検索における商品に関連する意図として Comparison, Support を加えた分類法を提案した [30]. Ashkan と Clarke はクエリを商用と非商用に分類した [31]. 前述したように、Su らは EC サイトにおける検索意図の分類法を動機、検索対象の特異性、検索戦略に基づき提案した [4]. Moe は EC サイトにおけるユーザのクリック行動に基づき意図を分類した [32]. Sondhi らは EC サイトのクエリについて、クリック率やクエリ長などを特徴量として5つのカテゴリにクラスタリングした [33].

これらの研究は検索行動を理解するために検索意図の分類や予測などを行っている.対して本研究では、購買に関連するWeb 検索クエリを調査するだけでなく、検索と購買の関係も分析する.また、予測対象も異なり、検索意図ではなく、購買満足度を予測することを目的とする.本研究では情報探索行動における検索の深さに特に興味があるため、既存のクエリ分類法の中でもSuらが提案した、情報の特異性に基づく分類法を採用する.

3 対象データ

本節では、本研究で用いたデータセットについて述べた後、 クエリの背後にある、商品に関連する検索意図を推定する方法 を示す.

3.1 データセット

本研究では商品購買前後の Web 検索行動と満足度の関係を分析するため、日本の大手商用サーチエンジンである Yahoo! 検索の Web 検索ログと、日本の大手 EC サイトである Yahoo! ショッピングの購買ログおよびレビューデータをサンプリング

大カテゴリ	カテゴリ
家電	扇風機,空気清浄機,炊飯器,電子レンジ・オーブンレンジ,エ アコン,冷蔵庫,洗濯機,掃除機,食器乾燥機,食洗機
オーディオ	イヤホン・ヘッドホン,スピーカー,キーボード,電子ピアノ
美容	ヘアアイロン, シェーバー, ドライヤー, アイロン, 女性用シェー バー
ガジェット	マウス, デジカメ, プリンタ, 一眼レフ, ビデオカメラ, Web カメラ, プロジェクタ
アウトドア	寝袋, テント, 電動自転車, シティサイクル, クロスバイク, ロードバイク

し用いる。これら 2 つのサービスは同じ会社が運営しているため,同一ユーザの検索行動と購買行動を容易に把握することができた。サーチエンジンを日常的に利用しているユーザを中心に調査したため,毎月 10 日以上検索を行ったユーザのログのみを使用した。Web 検索と購買のログは 2016 年から 2017 年にかけて取得したものを用いた。例えば 2017 年 12 月に商品を購買した顧客が 2018 年 12 月にレビューを書くなど,レビューは購買後に書かれたものを扱うため,2016 年から 2018 年にかけて取得したものを使用した。

表1に各データの情報を示す.表1では分析対象となるサービスや期間に加え,ユーザ数,クエリ数,平均クエリ長,平均セッション長を記載している.なお,これらの数値は母集団からのサンプルであるため,サービスの総ユーザ数,総クエリ数を表すものではない.セッションについては,クエリログ分析でよく使われる30分の閾値を設け求めた[34],[35].平均セッション長は,各セッションの平均クエリ数とした.購買商品のカテゴリは,表2のように5つの大カテゴリに分類した.これにより,大カテゴリごとの検索行動の差異を調べることができた.

レビューデータには 5 段階の点数とレビュー文が含まれる. 点数は必ずしも購買満足度を示すものではないため,我々が以前取り組んだ研究を参考に,次の 3 つの条件によってフィルターをかけた:(1) 購買前のレビューであること (2) 50 文字以上のレビューであること (3) 同じユーザによる既存のレビューと同一ではないこと [5], [6]. これらのフィルターを適用して得られた統計量を表 1 に示す.満足度についても同様に,点数が4 または 5 の購買を満足のいく購買とし,それ以外を不満のある購買とした 1 .

3.2 検索意図の推定

Web 検索行動を特徴付けるため、我々が以前取り組んだ研究と同様にして、各クエリの検索意図が Target Finding (TF)、Decision Making (DM)、Other のいずれであると考えられるか推定した [5]、[6]. 具体的には、検索意図の分類法において、本研究に最も関連する Su らの研究は Target Finding (TF)、Decision Making (DM)、Exploration (EP)の検索意図を考案している [4]. この分類法は、(1) 商品検索における情報探索行

^{1:}単純化のため,(D)SAT 購買と (D)SAT ユーザは 1 対 1 に対応しないが,本研究では同じ意味で使用している(例: 2 つの商品を購買したユーザが,1 つ目の購買には満足するが 2 つ目の購買には満足しない).

表 3 弱教師あり学習アプローチの検証

大カテゴリ	正解率	F1 値
家電	0.92	0.67
オーディオ	0.93	0.53
美容	0.97	0.40
ガジェット	0.89	0.48
アウトドア	0.85	0.60
平均	0.91	0.53

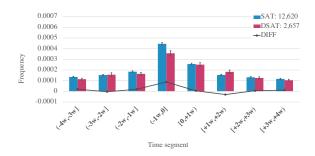
動を分析するために設計されており、本研究の分析に適していると考えられ、(2) 前述したように、検索目的の具体性に基づくものがほとんどであった.本研究では、情報探索の異なる段階におけるこれらの検索意図の量は、購買後の満足度の違いによって有意に異なる可能性があると仮説を立てた.なお、EPは他の意図とは大きく異なり、Web クエリから特定することは困難である.そこで、本研究では TF と DM のみを用い、TF にも DM にも当てはまらないクエリの検索意図を Other とした.

Su らの分析では、各セッション (合計 1,713 セッション)を手動でアノテーションしているが、Su らの手法をデータセットにある 4,500 万以上の Web クエリに適用することは難しい、そのため、クエリごとに検索意図を推定し、特定の商品名または特定のブランド名での商品検索意図を TF、商品タイプのみでの商品検索意図を DM と再定義しシステマティックに特定した、この際、DM を意味するクエリは多種多様であるため、DM の意図を推定することは容易ではない。したがって、ルールベースによる意図の推定の結果を取り入れ、大カテゴリごとに弱教師あり学習を用いて DM を特定した。

検索意図の推定結果を評価するため,各クエリを手動で評価し,その背後にある検索意図を推定した.各大カテゴリに対して 100 個のクエリをランダムに抽出し,合計 500 個のクエリを抽出した.アノテーションには,コンピュータサイエンス専攻の学生 3 名が参加した.各クエリに 2 人のアノテータが割り当てられ,クエリの背後にある意図が TF または DM と推定されるか否かを判定した.2 人の結果が異なる場合,3 人目の学生が多数決で意図を推定した.各アノテータが判定したクエリ集合が異なるため,Free-marginal Multirater Fleiss' Kappa [36]で評価者間の一致度を算出したところ, $\kappa=0.808$ と高い一致度が得られた.各大カテゴリにおける意図推定の評価結果を表3に示す.全体の正解率は十分高く,平均 F1 スコアは中程度と考えられる.

4 分 析

本節では、3節で推定した検索意図の時間分布を分析することで、RQ1、RQ2、RQ3への回答を試みる。検索ログをクエリから購買までの相対的な時間差に基づき 10 グループに分類した。各検索意図のクエリ頻度を総クエリ頻度で正規化した相対頻度を、各ユーザと時区間について計算、すなわち、 $\frac{f_{u,t,i}}{f_{u,\cdot,\cdot}}$ とした。 $f_{u,t,i}$ は時区間 t におけるユーザ u の検索意図 i のクエリ頻度を、 $f_{u,\cdot,\cdot}$ はすべての時区間におけるユーザ u のクエリ頻度を表す。



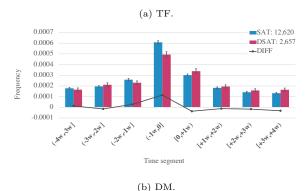


図 1 各検索意図の時間分布

4.1 満足したユーザと不満足なユーザで Web 検索行動は異なるのか?

図 1a は、SAT (満足) ユーザと DSAT (不満足な) ユーザの TF クエリの時間変化を示している。同様に、図 1b は、DM クエリの時間変化を示している。横軸はクエリから購買までの相対的な時間差を表す。0 が購買時を表すとする。((s)w,(s+1)w] (or [(s)w,(s+1)w)) は,購買前/後のs 週からs+1 週前/後の間にクエリを発行した時区間を示す。例えば,[0,+1w) は購買後 1 週間以内の時区間を表す。なお,4 週間前/後の時間区分の相対頻度はスケールの違いのため省略した。縦軸は各検索意図の相対頻度を表す。青と赤の棒グラフはそれぞれ SAT と DSAT の相対頻度を表し、比較のため,黒の線はその差を表している。また,図の右上に SAT と DSAT のユーザ数を示す。

図 1a と図 1b から,SAT ユーザの方が DSAT ユーザよりも 購買前に TF クエリで多く検索しており,特に購買前 1 週間以内ではその傾向が見られることが分かる。また,(-1w,0] における TF クエリと DM クエリにおける SAT と DSAT の差を比較すると,DM クエリの方がわずかに差が大きいように見られる。その他の時区間では,TF クエリの頻度は SAT と DSAT の間であまり差がなく,DM の頻度は SAT の方が低いことがわかる。これらの結果から,SAT ユーザは購買前 1 週間以内に商品に関連する情報を検索する頻度が高く,DM クエリの頻度が高いことから,より幅広い情報を重視していることが示唆される。相対頻度が正規分布に従わない可能性があるため,各時区間の各意図について,SAT ユーザと DSAT ユーザの差について Mann-Whitney U 検定を実施した。Holm-Bonferroni 法で有意水準を調整し,TF, DM ともに (-1w,0] に統計的に有意な差があることがわかった $(\alpha=0.05)$.

そして、SAT ユーザは DSAT ユーザと同じ範囲の情報を単純 に頻繁に探索するのではなく、より広い範囲の情報を探している

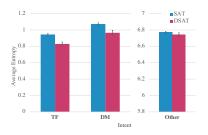


図 2 クエリ語エントロピーの平均

という仮説を立てた.この仮説を検証するために,購買前の時区間に着目し,検索クエリで使用される単語の多様性をクエリ語エントロピーで比較した. $Q_{u,i}$ を,購買の4週間前までに検索意図iを持つユーザuが発行したクエリの集合,Vを単語の集合全体とする.検索意図iを持つユーザuに対する単語vの確率 $p_{u,i}(t)$ は,次のように定義される: $p_{u,i}(v)=c_{u,i}(v)/\sum_{v'\in V}c_{u,i}(v')$.ここで, $c_{u,i}(v)$ は $Q_{u,i}$ における単語vの頻度を表す.そして, $e_{u,i}=-\sum_{v\in V}p_{u,i}(v)\log_2(p_{u,i}(v))$ によってクエリ語エントロピーを求めることができる.

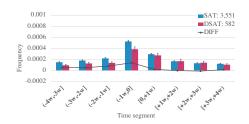
図 2 は、SAT ユーザと DSAT ユーザの検索意図 (横軸) ごとの平均クエリ語エントロピー (縦軸) を示している。参考のために Other クエリについても平均クエリ語エントロピーを計算しているが、ここでは SAT ユーザと DSAT ユーザの間に大きな差は見られない.一方、TF と DM の両クエリにおいて、DSAT ユーザよりも SAT ユーザの方が多様なクエリ語を使用していることがわかる.2 元配置分散分析の結果、満足度の効果は統計的に有意であった (F(1,45827)=25.6,p<0.05). Tukey HSD では、SAT と DSAT のユーザ間で統計的に有意な差があることが示された $(\alpha=0.05)$. さらに、DM クエリはTF クエリよりも多様な用語を含むことが観察される.

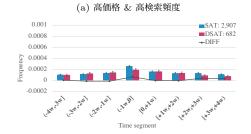
要約 RQ1 に対して、SAT ユーザは DSAT ユーザよりも購買前 1 週間以内に、より広い商品に関連する情報を検索する頻度が高いと結論付けることができる。Punj と Staelin は、外部探索量の増加によって消費者が小売価格よりも安い価格で商品を購買することができ、そのコスト削減が顧客満足度に正の影響を与えることを発見した [17]. 本小節の知見は、外部探索量と顧客満足度の間接的関係に関する彼らの研究を支持する.

4.2 商品やユーザの特性が変われば,満足したユーザと不満 足なユーザで Web 検索行動は異なるのか?

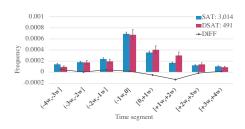
本小節では、購買商品を価格、顧客を検索頻度で特徴付けることで、RQ2と RQ3を検証する。商品価格はマーケティングの研究において商品を特徴付けるためによく使われ (e.g., [14])、検索頻度は情報検索の研究において検索の専門性の代わりとして使われる (e.g., [37])。商品は高価格群 (商品価格が中央値より高い群)と低価格群に分け、顧客は高検索頻度群 (Web での情報探索に慣れている顧客と考えられる、検索頻度の中央値以上の検索を行った群)と低検索頻度群に分けた。商品価格と検索頻度が購買後の満足度に相互作用を及ぼす可能性があるため、可能な組み合わせで調査を行った。

各価格と検索頻度の組み合わせにおける, TF クエリと DM クエリの時間分布をそれぞれ図 3 と図 4 に示す. 図 3 から,











(d) 低価格 & 低検索頻度

図 3 商品価格と検索頻度の各組み合わせにおける TF クエリの時間 分布

SAT ユーザは DSAT ユーザよりも、商品価格と検索頻度が高い場合にのみ、購買前に TF クエリで頻繁に検索していることがわかる (3a). さらに、検索頻度が低く、かつ高価な商品を購入した場合、DSAT ユーザは SAT ユーザよりも購買後 2 週間以内の TF クエリの検索頻度が高いことがわかる (3c). その他の場合では、SAT ユーザと DSAT ユーザの間にあまり差がないため、図 1a の傾向は高価格群にのみ見られることが判明した. そこで、前小節と同様に Mann-Whitney U 検定を行い、Holm-Bonferroni 法で有意水準を調整した. その結果、図 3aの (-3w, -2w]と (-1w, 0] においてのみ有意差が認められた.

TF クエリで有意差が見られたことから、SAT ユーザは、特に Web 検索に慣れている場合や高価な商品を購買する場合、購買前に具体的な情報を頻繁に検索していることが示唆された.この傾向は、知覚リスクと顧客満足度の関係に関する既存の研究によって説明できる可能性がある. Dowling と Staelin は、知覚リスクには許容値があると主張した [38]. 彼らは、知覚リスクが許容値以上である場合、顧客は情報探索によって知覚さ



(d) 低価格 & 低検索頻度

図 4 商品価格と検索頻度の各組み合わせにおける DM クエリの時間 分布

れたリスクを下げようとすることを示した.一方,知覚リスク量が許容値を下回る場合には,知覚リスクは情報探索行動にほとんど影響を与えないことが示された.また,Ghotbabadiらは,知覚リスクは満足度と負の相関があり,顧客が満足するためには知覚リスクを下げることが重要であることを示した[39]. 最後に,Jacoby と Kaplan は知覚リスクの構成要素を紹介し,その1つとして financial risk があることを示した [40]. 本分析においては,顧客が高価な商品を購買しようとする場合,安価な商品を購買する場合と比較して,より高い financial risk を持つはずである.もし,顧客が Web 検索に慣れていれば,特定の商品について Web 検索を行い,知覚リスクの構成要素の1つである financial risk を低減させることで,知覚リスクを低減させることができる可能性がある.したがって,顧客は知覚リスクの低減に成功したことで,購買に満足できたと考えられる.

TF クエリとは対照的に、購買前の DM クエリは Web 検索の 頻度が低い SAT ユーザの投入が多いことが図 4 から見て取れ る. また、(-3w, -2w] と購買後では、検索頻度の低い DSAT ユーザが投入した DM クエリの量が多くなっていることが見て取れる (4d). 図 4d 中の (-3w, -2w], [+3w, +4w) でのみ有意差が見られた.

このような有意差は説明が難しいが、顧客の再購買のシグナルと考えることができる。顧客は購買した商品に不満があり、かつ商品価格が安い場合、購買した商品の代替品を探す方向に向かうかもしれない。一方、高価な商品を購買した後では、予算の関係で代替品を購入することが難しくなる。

要約 RQ2, RQ3 に対して、顧客の検索行動は特に Web 検索に比較的慣れていて、比較的高価な商品を購買する場合、TF クエリの頻度が SAT ユーザと DSAT ユーザで異なっていたと結論付けられた。 SAT ユーザは、 DSAT ユーザよりも購買前に多くの TF クエリで検索していた。 さらに、検索頻度が低く、安価な商品を購買するユーザにおいては、購買後に DSAT ユーザの方が SAT ユーザより多くの DM クエリを投入していた。

4.3 クエリ語の分析

本小節では、TF および DM クエリで使用される語と満足度 との関係を分析する. これまでの分析で、SAT ユーザと DSAT ユーザの間では、特に購買前の TF と DM のクエリの頻度に 違いがあることが判明している. この知見に基づき、購買前に 投入された検索クエリに着目し、どのようなクエリ語が使用さ れているかを分析した. まず, 各ユーザの TF クエリと DM ク エリに含まれる単語 (fastText [41]) の埋め込みを平均化し、各 ユーザを表すベクトルを求めた. SAT ユーザと DSAT ユーザ で検索に用いる単語が異なるかを明らかにするため、SAT ユー ザの各購買を表すベクトルと、DSAT ユーザの各購買を表すべ クトルを比較した. 加えて, ユーザが検索に用いる単語は多様 であると考えられるため、各購買を表すベクトルを代表的なク ラスタリング手法である k-means 法でクラスタリングし、各ク ラスタでどのような特徴が見られるか比較した. k の値を変え て適切なクラスタ数を探索し、最終的にk=5が最も情報量の 多いクラスタを生成できることを見出した.

その後,購買に関する単語は多様であるため,第 1 著者が人手で確認し抽出,購買において似た意味を持つ語でグループ化した.

図5は、各クラスタに含まれる単語群の頻度をヒートマップで表したものである。ヒートマップの各行は単語群を表し、各列はクラスタを表す。各セルは、各クラスタにおける累積の単語頻度を正規化した値を表している。図6は、各クラスタにおける購買と検索に関する指標の平均値を示している。ヒートマップの各行は、上から順に、SATユーザの割合、TFクエリでの平均検索回数、DMクエリでの平均検索回数、購買商品の平均価格を表している。各列はクラスタを表す。各セルの色は、クラスタ間の順位を表している。

Cluster 1 は、平均的なユーザ群と考えられ、他との顕著な差は見られない. Cluster 2 は、EC サイトを直接訪問し (Term group 2、"(EC サイト X)" または"(小売店 X)" は特定のサイト名を表す)、色や価格などの商品の特徴を検索する

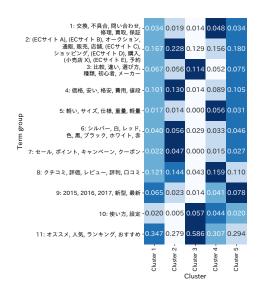


図 5 各クラスタの各単語群における検索頻度

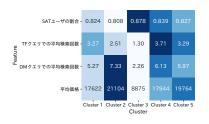


図 6 各クラスタの特徴

ユーザを含む.また、DM クエリでの検索を頻繁に行い、高価な商品を購買し、購買した商品に不満を持つ可能性が高い.一方、Cluster 3 は、提案やおすすめを求めるユーザ群である.検索回数は少ないが、購買した商品に満足する可能性が高い.Cluster 4 は、TF や DM クエリでレビューや商品の詳細情報を頻繁に検索する、満足度の高いユーザ群である.Cluster 5 は、特徴づけが難しいが、年に関するクエリで検索しているユーザを含む.

要約 本小節の分析は前小節に比べ探索的であるが、ユーザのクラスタリングにより、他者の意見(比較、違い、選び方、初心者、おすすめ、ランキング、評価、レビュー)を求めている顧客は、購買に満足する可能性が高いことがわかった.一方、色や価格など特定の機能を探し、複数のECサイトを訪問したユーザは、満足のいく購買に至らない可能性が高い.クエリ語と購買満足度に因果関係があると強く主張することはできないが、他の調査結果と同様に、商品に関連する検索で使用される語の種類とその特徴が見て取れた.

5 満足度予測

本節では、ユーザの時区間ごとの検索回数と Web 検索クエリに含まれる単語をそれぞれ用いて購買満足度の予測を行うことで、RQ4 を検証する.本研究では問題を簡略化するため、4節で想定したユーザの満足度を予測する2値分類問題に取り組む.また、表1にあるように、82.6%が SAT 購買に分類される.偏った予測を避け、バランスのとれたデータセットを作成

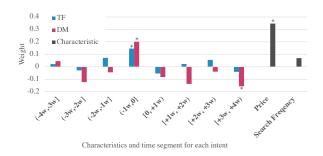


図 7 満足度予測モデルの重み. アスタリスクはその要因が統計的に 有意であることを示す.

するために、SAT 購買をアンダーサンプリングした.

5.1 検索意図の時間分布による予測

4節でSATとDSATのユーザ間でいくつかの違いが見られ たため、まず TF と DM クエリの時間分布に基づく予測を試み た. 購買前後の8時区間で投入されたTFとDMクエリの正 規化された頻度で、各ユーザの8次元ベクトルを表す. さらに、 価格と検索頻度の水準を表すバイナリの特徴量を含めた. 満足 度予測には、合計 18 個 (= $8 \times 2 + 2$) の特徴量を使用した。分 類モデルとしては、ランダムフォレストとロジスティック回帰 を用いた. ランダムフォレストは、木の本数を1から20本、木 の深さを1から5本とすることで、複数のパラメータを調整す ることができる. ロジスティック回帰では, 正則化の強さの係数 を 10 倍刻みで 10^{-2} から 10^{5} の範囲でハイパーパラメータを設 定する. 学習とテストには、3-fold nested cross validation を 用いた. 最適なパラメータはグリッドサーチによって決定した. その結果、ランダムフォレストは正解率が 0.5514、ロジス ティック回帰は正解率が 0.5591 を達成した. これは、予測タス クが非常に難しいにも関わらず, 予測モデルがランダムな予測 (0.5) を上回っていることを示している。また、ロジスティック 回帰モデルによる各特徴の重みを図7に示す. 横軸はモデルで 使用した特徴量、縦軸は各特徴量の重みを表している. 購買前 1 週間以内の DM と TF の頻度, [+3w, +4w) の DM の頻度, 価格特徴が統計的に有意な要因であることがわかった. この傾 向は4節で得た、満足したユーザはそうでないユーザと比べ、 特に購買前1週間以内では頻度が高い傾向を捉えていると考え られる. また、購買後の DM クエリ、価格において重みの絶対 値が大きい傾向も見られる. これらの傾向は、価格が中央値よ りも高い場合において頻度の差が見られることや, DM クエリ においては、価格とユーザの検索回数が中央値以下の場合に、 満足したユーザはそうでないユーザと比べ購買後の頻度が低い 傾向を捉えていると考えられる. これらの傾向について、GLM (Generalized Linear Model) で検証したところ、4 変数の影響 がいずれも統計的に有意だった. 具体的には, 購買前1週間 以内の TF クエリ (z = 2.466, p < 0.05, 95% CI[0.080, 0.697])購買前 1 週間以内の DM クエリ (z=2.978,p<0.05,95% CI[0.141,0.686]), 購買後3週間以降4週間以内の DM クエリ (z = -1.983, p < 0.05, 95% CI[-1.093, -0.007]),価格 (z = 7.807, p < 0.05, 95% CI[0.332, 0.554]) で統計的に有

意だった.しかし,購買前1週間以外の購買前における頻度や, ユーザの検索回数については有意な差は見られなかった. これ らの特徴を捉えられなかった原因として、アンダーサンプリン グをした結果、十分な学習が行えず満足するユーザに共通する 特徴を見つけ出せなかったことが推測される.

5.2 クエリ語による予測

クエリ語による予測では、bag-of-words、クエリ埋め込み、 BERT の3種類の手法において、購買前4週間のTFクエリ と DM クエリを空白で分割して得た語を用いて満足度の予測を 行った. bag-of-words による予測では、クエリを分割して得た 語から bag-of-words を作成し,入力として満足度の予測を行っ た. この際、購買に関する語は多様であるうえ、一部のユーザ のみが極端に多く用いている語に影響される可能性がある. そ こで、商品を購買したユーザが共通して検索に用いている語に 限定するため、10人以上のユーザが用いていた語に限定して bag-of-words を作成した. 分類器には前小節でより高い正解 率を記録したロジスティック回帰を用いた. クエリ埋め込みに よる予測では、予測に用いるクエリについて、4.3節と同様に Grave らの fastText による学習済み単語の埋め込みを平均し、 各購買を表すベクトルを作成した[41]. 加えて, 前小節と同様 に、購買した商品の価格が中央値よりも高いかどうかを表すバ イナリ値と、分析期間内での検索回数が中央値よりも高いユー ザかどうかを表すバイナリ値を各ベクトルに連結させた. 以上 の手順で得たベクトルを入力としてクエリ埋め込みによる満足 度の予測を行った. BERT による予測では、Hugging Face に て公開されている日本語の事前学習済みモデル 2を用いて、ク エリの時系列を考慮した予測を行った. BERT に代表される事 前学習済み言語モデルでは、解きたいタスクと類似したデータ を用いて追加学習を行うことで、予測精度が向上することがわ かっている [42]. そのため、事前学習済みの BERT モデルに対 し、購買に関するクエリで追加学習を行った. 追加学習には、 満足度予測の対象となるユーザおよびクエリと条件を揃えるた め、分析対象期間に商品を購買しており、かつ毎月 10 日以上 検索を行ったユーザの TF クエリまたは DM クエリを用いた. これらのクエリを時系列順に並べ、クエリ間は SEP トークン で区切って入力として教師なし学習を行った. なお、BERT の 追加学習にあたっては、Next Sentence Prediction タスクはあ まり有用でないことが知られている[43]. そのため、Masked LM タスクのみを解く追加学習を行った. ファインチューニン グについても同様の形式で TF クエリまたは DM クエリを入力 として行い, データは訓練データ, 検証データ, テストデータ の比が順に 8:1:1 となるように分割した. 学習率は 10^{-6} , 10^{-5} , 10^{-4} , バッチサイズは 16,32 の候補の組合せからグリッドサー チを行い決定した.

表4に、意図の頻度を用いた場合とクエリ語を用いた場合の 満足度予測の正解率を示す.表4では、予測に用いた各手法に ついて、用いた時区間および特性ごとに正解率が示されている.

表 4 各モデルでの正解率

表 4 各モデルでの正解率		表 5	5 BERT の時区間ごとの正	
モデル	正解率		時区間	正解率
意図の頻度	0.533		(-1w, 0]	0.529
意図の頻度+購買後	0.531		(-2w, 0]	0.524
意図の頻度+購買後+特性	0.559		(-3w, 0]	0.537
bag-of-words	0.530		(-4w, 0]	0.548
クエリ埋め込み	0.529		(-4w, +1w)	0.542
クエリ埋め込み+特性	0.559		(-4w, +2w)	0.529
BERT	0.548		(-4w, +3w)	0.535
BERT+特性	0.533		(-4w, +4w)	0.541

表中の"+購買後"は購買前4週間に加え,購買後4週間のク エリも用いて予測した場合を表している. "+特性" は商品特性 やユーザ特性を表すバイナリ値も用いた場合を表している.表 4から、特性を用いない場合においては、BERT による予測が 最も高い正解率を記録していることがわかる. したがって、ク エリ語を用いた手法において、モデルの複雑性や時系列が予測 に比較的有用であると考えられる. 一方, 特性を用いた場合に おいては、意図の頻度を用いた手法とクエリの埋め込みを用い た手法において、BERT による予測の正解率を上回っているこ とがわかる. この傾向から, 前小節で検証し有意差が見られた 特性である価格が、意図の頻度だけでなくクエリ語による予測 においても大きく寄与していると考えられる. 加えて、BERT による予測では特性を捉えた予測が行えていないことが推測で きる. その原因としては、本実験での予測に用いたデータ数が、 BERT がクエリ語と特性の組合せを学習するために十分な量で はなかったことが考えられる. また, 意図の頻度による予測で は、購買前4週間のクエリで予測した場合よりも、購買後4週 間のクエリも用いた場合の方が正解率が低くなっている. した がって, 意図の頻度による予測では, 購買後のクエリは予測に 寄与するとはいえないと考えられる.

表5に、BERTでの予測において、用いた時区間を変えた場 合の満足度予測の正解率を示す.表中の時区間はどの時区間に 入力された TF クエリと DM クエリを用いるかを表しており, 例えば、(-1w, 0] は購買前1週間以内から購買時、(-4w, +2w)は購買前4週間から購買後2週間以内の時区間で入力されたク エリを用いて予測したことを表している.表5から、時区間が 長くなるほど正解率が上がるとは必ずしもいえないことがわか る. したがって、特定の時区間のクエリ語や、時区間の長さが 予測に効いているといえないと考えられる.

6 議論とまとめ

本研究では、EC サイトにおける商品購買の前後に行われる Web 検索行動と購買満足度の関係分析を行った. 分析や予測 を通して得た知見をもとにインプリケーションを示す. 既存研 究では、知覚リスクが許容値以上である商品を購買する際に は、商品に関する検索によって知覚リスクを下げることがユー ザの満足に結びつくと考えられたが、本研究でもユーザ行動ロ グからこの知見を支持する結果が示された[38],[39],[40]. した がって, ユーザが商品の購買を検討していると思われる場合, 特にそれがよく検索するユーザかつ高価な商品である場合、そ のユーザが商品に関してよく検索することで知覚リスクが減少

 $^{2 \}verb|:https://huggingface.co/cl-tohoku/bert-base-japanese-whole-wor$ d-masking

するように促す仕組みが必要だと考えられる. また, 購買に満足したユーザはそうでないユーザと比べ, 購買前に商品に関する検索の頻度が高く, 多様な語で商品に関する検索をしていることも推測された. 特に購買前1週間以内の商品に関する検索の頻度は満足度予測にも有効だった. したがって, 購買前, 特に購買直前において, 多様な検索を促す Web 検索アルゴリズムやインターフェースが必要だと考えられる. 加えて, 特に購買前1週間以内の商品に関する検索の頻度は満足度予測にも有効だったことから, 購買直前において, 多様な検索を促す Web 検索アルゴリズムやインターフェースが必要だと考えられる.

本研究では Web 情報探索行動と購買満足度の関係を調査し た. Web 情報探索行動の記録として Web 検索ログを, 購買満 足度の評価として EC サイトにおける商品レビューの点数を用 いて分析を行った.分析にあたっては、EC サイトで商品を購 買したユーザを特定し、その検索意図に基づいて Web 検索ク エリを分類, 購買満足度が異なるユーザの時間分布を調査した. 加えて、用いたデータを商品特性、ユーザ特性に着目して掘り 下げ、Web 検索の違いを調査した他、顧客が検索に用いるク エリ語についても分析した. 分析の結果, SAT (満足した) 顧 客と DSAT (不満足な) 顧客, さらには顧客のタイプや商品に よって、Web 検索行動が異なることが明らかになった。特に、 (1) 購買前 1 週間以内に, SAT ユーザは DSAT ユーザに比べ, より幅広く商品に関連する情報を検索する頻度が高いことがわ かった. (2) SAT ユーザは DSAT ユーザよりも購買前に TF ク エリで多く検索しており、特に Web 検索に比較的慣れている 場合や比較的高価な商品を購買する場合, DSAT ユーザよりも 多く具体的なクエリで検索していた. (3) 他人の意見を求めて いる顧客は, 購買時に満足する可能性が高かった. また, 購買 前後の Web 検索行動から、顧客の購買満足度を予測する問題 にも取り組んだ. この試みにより, 顧客が行った Web 検索の 種類が異なれば、購買満足度も異なる可能性があることが実証 された.

今後の課題として,より詳細な検索意図での分類や,購買満足度の予測正解率の向上などが挙げられる.

謝辞 本研究は JSPS 科学研究費助成事業 JP22H03905, JP21H03775, JP21H03774, JP18H03244 による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] PwC. December 2021 global consumer insights pulse survey, 2021. https://www.pwc.com/gx/en/industries/consumer-markets/consumer-insights-survey/archive/consumer-insights-survey-december-2021.html, (accessed 2022-11-28).
- [2] Marsha L. Richins and Peter H. Bloch. Post-purchase product satisfaction: Incorporating the effects of involvement and time. *Journal of Business Research*, Vol. 23, No. 2, pp. 145–158, 1991.
- [3] Rodoula Tsiotsou. The role of perceived product quality and overall satisfaction on purchase intentions. *International* journal of consumer studies, Vol. 30, No. 2, pp. 207–217, 2006.
- [4] Ning Su, Jiyin He, Yiqun Liu, Min Zhang, and Shaoping

- Ma. User intent, behaviour, and perceived satisfaction in product search. In WSDM, pp. 547–555, 2018.
- [5] 柳田雄輝, 加藤誠, 河田友香, 山本岳洋, 大島裕明, 藤田澄男. 検索行動に基づく購買満足度の関係分析. 第14回データ工学と情報マネジメントに関するフォーラム, 2022.
- [6] 柳田雄輝, 加藤誠, 河田友香, 山本岳洋, 大島裕明, 藤田澄男. 検索 行動に基づく購買満足度の関係分析. データドリブンスタディー ズ, Vol. 1, No. 3, 2023.
- [7] Ismail Razak, Nazief Nirwanto, and Boge Triatmanto. The impact of product quality and price on customer satisfaction with the mediator of customer value. *IISTE: Journal* of Marketing and Consumer Research, Vol. 30, pp. 59–68, 2016.
- [8] Roger Hallowell. The relationships of customer satisfaction, customer loyalty, and profitability: an empirical study. *International journal of service industry management*, Vol. 7, No. 4, pp. 27–42, 1996.
- [9] Anders Gustafsson, Michael D. Johnson, and Inger Roos. The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention. *Journal of marketing*, Vol. 69, No. 4, pp. 210–218, 2005.
- [10] Ruth Mugge, Hendrik N. J. Schifferstein, and Jan P. L. Schoormans. Product attachment and satisfaction: understanding consumers' post-purchase behavior. *Journal of consumer Marketing*, Vol. 27, No. 3, pp. 271–282, 2010.
- [11] Todd A. Mooradian and James M. Olver. "i can't get no satisfaction:" the impact of personality and emotion on postpurchase processes. *Psychology & Marketing*, Vol. 14, No. 4, pp. 379–393, 1997.
- [12] Magnus Söderlund. Customer familiarity and its effects on satisfaction and behavioral intentions. Psychology & Marketing, Vol. 19, No. 10, pp. 861–879, 2002.
- [13] Richard N. Cardozo. An experimental study of customer effort, expectation, and satisfaction. *Journal of marketing* research, Vol. 2, No. 3, pp. 244–249, 1965.
- [14] Yuren Wang, Xin Lu, and Yuejin Tan. Impact of product attributes on customer satisfaction: An analysis of online reviews for washing machines. *Electronic Commerce Re*search and Applications, Vol. 29, pp. 1–11, 2018.
- [15] Howard Beales, Michael B. Mazis, Steven C. Salop, and Richard Staelin. Consumer search and public policy. *Journal of Consumer Research*, Vol. 8, No. 1, pp. 11–22, 1981.
- [16] Janet R. McColl-Kennedy and Richard E. Fetter. An empirical examination of the involvement to external search relationship in services marketing. *Journal of services marketing*, Vol. 15, No. 2, pp. 82–98, 2001.
- [17] Girish N. Punj and Richard Staelin. A model of consumer information search behavior for new automobiles. *Journal* of consumer research, Vol. 9, No. 4, pp. 366–380, 1983.
- [18] Sharon E. Beatty and Scott M. Smith. External search effort: An investigation across several product categories. *Journal of consumer research*, Vol. 14, No. 1, pp. 83–95, 1987.
- [19] Narasimhan Srinivasan and Brian T. Ratchford. An empirical test of a model of external search for automobiles. *Journal of Consumer research*, Vol. 18, No. 2, pp. 233–242, 1991.
- [20] Sharon Hirsch, Ido Guy, Alexander Nus, Arnon Dagan, and Oren Kurland. Query reformulation in e-commerce search. In SIGIR, pp. 1319–1328, 2020.
- [21] Rohan Kumar, Mohit Kumar, Neil Shah, and Christos Faloutsos. Did we get it right? predicting query performance in e-commerce search. In eCom@SIGIR, 2018.
- $[22] \quad \text{Yangyang Guo, Zhiyong Cheng, Liqiang Nie, Xin-Shun Xu,} \\ \text{and Mohan Kankanhalli. Multi-modal preference modeling} \\ \text{for product search. In MM, pp. 1865–1873, 2018.}$
- [23] Bernard J. Jansen and Simone Schuster. Bidding on the buying funnel for sponsored search and keyword advertis-

- ing. Journal of Electronic Commerce Research, Vol. 12, No. 1, pp. 1–18, 2011.
- [24] Beibei Li, Anindya Ghose, and Panagiotis G. Ipeirotis. Towards a theory model for product search. In WWW, pp. 327–336, 2011.
- [25] Qianyun Zhang, Shawndra Hill, and David Rothschild. Post purchase search engine marketing. In WWW, pp. 663–670, 2018
- [26] Adrian Boteanu, Emily Dutile, Adam Kiezun, and Shay Artzi. Subjective search intent predictions using customer reviews. In CHIIR, pp. 303–307, 2020.
- [27] Long Guo, Lifeng Hua, Rongfei Jia, Binqiang Zhao, Xiaobo Wang, and Bin Cui. Buying or browsing?: Predicting realtime purchasing intent using attention-based deep network with multiple behavior. In KDD, pp. 1984–1992, 2019.
- [28] Andrei Broder. A taxonomy of web search. In ACM SIGIR forum, Vol. 36, pp. 3–10, 2002.
- [29] Bernard J. Jansen, Danielle L. Booth, and Amanda Spink. Determining the informational, navigational, and transactional intent of web queries. *Information Processing & Management*, Vol. 44, No. 3, pp. 1251–1266, 2008.
- [30] Nikitha Rao, Chetan Bansal, Subhabrata Mukherjee, and Chandra Maddila. Product insights: Analyzing product intents in web search. In CIKM, pp. 2189–2192, 2020.
- [31] Azin Ashkan and Charles L. A. Clarke. Characterizing commercial intent. In CIKM, pp. 67–76, 2009.
- [32] Wendy W. Moe. Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of consumer psychology*, Vol. 13, No. 1-2, pp. 29–39, 2003.
- [33] Parikshit Sondhi, Mohit Sharma, Pranam Kolari, and ChengXiang Zhai. A taxonomy of queries for e-commerce search. In SIGIR, pp. 1245–1248, 2018.
- [34] Benjamin Piwowarski and Hugo Zaragoza. Predictive user click models based on click-through history. In CIKM, pp. 175–182, 2007.
- [35] Jaime Teevan, Eytan Adar, Rosie Jones, and Michael A. S. Potts. Information re-retrieval: Repeat queries in yahoo's logs. In SIGIR, pp. 151–158, 2007.
- [36] Justus J. Randolph. Free-marginal multirater kappa (multirater k [free]): An alternative to fleiss' fixed-marginal multirater kappa. Online submission, 2005.
- [37] Makoto P. Kato, Takehiro Yamamoto, Hiroaki Ohshima, and Katsumi Tanaka. Investigating users' query formulations for cognitive search intents. In SIGIR, pp. 577–586, 2014.
- [38] Grahame R. Dowling and Richard Staelin. A model of perceived risk and intended risk-handling activity. *Journal of consumer research*, Vol. 21, No. 1, pp. 119–134, 1994.
- [39] Ali Ramezani Ghotbabadi, Setareh Feiz, and Rohaizat Baharun. The relationship of customer perceived risk and customer satisfaction. *Mediterranean Journal of Social Sciences*, Vol. 7, No. 1 S1, pp. 161–173, 2016.
- [40] Jacob Jacoby and Leon B. Kaplan. The components of perceived risk. ACR special volumes, 1972.
- [41] Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. Learning word vectors for 157 languages. In *LREC*, pp. 3483–3487, 2018.
- [42] Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A Smith. Don't stop pretraining: Adapt language models to domains and tasks. In ACL, pp. 8342–8360, 2020.
- [43] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A robustly optimized BERT pretraining approach. In ICLR, 2020.