

# ECサイトでのカメラ購買前後における Web 検索クエリの時系列分析

河田 友香<sup>†</sup> 山本 岳洋<sup>†,††</sup> 大島 裕明<sup>†,††</sup> 柳田 雄輝<sup>†††</sup> 加藤 誠<sup>†††</sup>  
藤田 澄男<sup>††††</sup>

<sup>†</sup> 兵庫県立大学 社会情報科学部 〒 651-2197 神戸市西区学園西町 8-2-1

<sup>††</sup> 兵庫県立大学 大学院情報科学研究科 〒 650-0047 兵庫県神戸市中央区港島南町 7-1-28

<sup>†††</sup> 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

<sup>††††</sup> 筑波大学 図書館情報メディア系 / JST さきがけ 〒 305-8550 茨城県つくば市春日 1-2

<sup>†††††</sup> ヤフー株式会社 〒 102-8282 東京都千代田区紀尾井町 1-3

E-mail: †{fa19c021,t.yamamoto}@sis.u-hyogo.ac.jp, ††ohshima@ai.u-hyogo.ac.jp,

†††††ynagi2@klis.tsukuba.ac.jp, †††††mpkato@acm.org, †††††sufujita@yahoo-corp.jp

**あらまし** 本研究では、EC サイトでカメラを購入したユーザを対象に、そのユーザの Web 検索クエリ中に含まれる語が購買前から購買後にかけて時間とともにどのように変化するかを調査した。まず、検索回数や検索される語の分析を行い、購買 48 時間前から検索回数が増え始めること、購買 24 時間前から購買直前にかけてはメーカー名、シリーズ名、型番に関する語が多くなることを明らかにした。また、カメラの購買前後に特徴的な 200 語を検索によって求める情報の違いから分類し、「付属品・補完品」、「配達」、「使用方法」といった 16 種類を明らかにした。得られた分類を用いて、どのような種類の語が購買前、購買後のいつごろ多くクエリに用いられるのかを分析した。その結果、購買後に関しては、購買直後から購買 24 時間後には付属品や補完品、購買 24 時間後から購買 48 時間後には配達、購買 48 時間後から購買 72 時間後には使用方法に関する語を含んだ検索が多くなることが分かった。

**キーワード** 情報検索, 購買行動, ユーザ行動分析

## 1 はじめに

オンラインショッピングにおいて Web 検索は重要な過程である。広告および市場調査会社である Survata は、オンラインショッピングを利用するユーザの 49% は特定商品の購買を決定していないときに Web 検索を行っており、また、36% のユーザは最初に Web 検索を行ってから商品の購買に至ると報告している [8]。このようにオンラインショッピングが重要な現在において、Web 検索は購買の判断をする重要な過程である。

また、商品購買に関する Web 検索においては、消費者の購買行動プロセスの理解が重要となる。ユーザの商品購買における興味や関心の状態に応じてクエリは変化していく [5]。Web 検索クエリを用いてユーザの購買行動への理解を深めることにより、広告や検索結果の改善、さらに Web 検索エンジンにおけるユーザ満足度の向上を目指すことができると考えられる [1] [10]。さらに、Web 検索クエリから商品購買におけるユーザの状態を理解することができれば、ユーザがいつごろ商品を購入するか [14] や商品を既に購買しているかといった予測も可能になると考えられる。

購買行動プロセスを理解するための一助となるのが Web 検索クエリである。ユーザが商品を購入するために、どのようなクエリを用いて Web 検索を行うのかを具体的に考えてみる。たとえば、「canon カメラ 種類」、「canon kiss x10i 価格」と検索を行い、商品の購買に至ったユーザがいたとする。「canon カメ

ラ 種類」で検索した際には、canon というメーカー名は定まっている一方で、購買する商品ははっきりとは決めていないと考えられる。「canon kiss x10i 価格」と検索した際には、具体的に購買する商品を「canon kiss x10i」と決めたと考え、価格を調べていると考えられる。特定の商品についての検索のほうが、そのユーザによる商品の購買に時間的に近いと考えられる。このように、クエリの変化を見ることにより、ユーザの商品購買におけるプロセスを理解することができるだろう。

購買前後においても、Web 検索クエリは変化すると考えられる。しかし、商品の購買前後での Web 検索クエリや Web 検索でのユーザ行動を分析した研究はまだ少ない。特に、購買前後によく検索されるクエリについて分析した研究はあるものの [12]、購買に至るまで、また購買後に検索される語が時間とともにどう変化するかを詳細に分析した研究はなされていない。

そこで本研究では、カメラの購買に関してユーザの Web 検索クエリログ、特にクエリの時間的遷移に着目し、購買に至るまでや購買した後のクエリを分析し、以下の 3 点を明らかにすることに取り組む。

**RQ1:** カメラの購買に関する検索はいつごろ、どのような語でなされるか。

**RQ2:** カメラの購買前、購買後に関する検索において、クエリに使用される語にどのような種類があるか。

**RQ3:** RQ2 で分類した語の種類は時間に応じてどのように変化していくか。

本研究では、以上の 3 点について、Yahoo!検索が提供する

Web 検索ログと Yahoo!ショッピングにおける商品購買履歴を分析する。

本研究で明らかになったことを以下にまとめる。

- **RQ1** オッズ比を用いて特定の時間区間において頻出する特徴的な語はあるか分析を行った。購買 24 時間前から購買直前に、「xp120」や「a10」といった商品名に関する語が多くなることが明らかとなった。

- **RQ2** ロジスティック回帰を用いて、与えられたクエリ集合がカメラの购买前に用いられるか、購買後に用いられるかを分類する分類器を構築し、得られた分類器から购买前、購買後に特徴的に表れる語を 200 語抽出した。

得られた語を、どのような情報を求めるため利用されるのかという観点から人手で分類することで、「型番」、「評価」、「購買サイト・店舗」、「付属品・補完品」、「配達」、「使用方法」といった 16 種類を明らかにした。

- **RQ3** RQ2 での分類に属する語が、購入の前後のいつごろ、検索クエリに現れやすいかを分析した。購買 24 時間前から購買直前においては、「メーカー」や「シリーズ」、「型番」に分類された語を含むクエリが増加することを明らかにした。購買直後から購買 24 時間後には「付属品・補完品」、購買 24 時間後から購買 48 時間後においては「配達」、購買 48 時間後から購買 72 時間後においては「使用方法」に分類された語を含むクエリが増加すること明らかとなった。

## 2 関連研究

### 2.1 商品に関する検索意図の分析

EC サイト上での検索や行動を用いた研究がなされている。Boteanu らは Amazon の消費者のレビュー文とクエリなどの購買に関するシグナルを用いて意図の予測を行った [2]。Su らは、アンケートとクエリログを用いて、Motivation, Target Specificity, Search Strategy という 3 つの観点からユーザの検索意図を分類し、Target Finding, Decision Making, Exploration という 3 種類に分類した [11]。Sondhi らの研究では EC サイト上の検索クエリから検索行動を特徴量とし、クラスタリングを行うことでクエリを 5 種類に分類した [10]。Manchanda らの研究では EC サイト上の検索から検索意図を定義し、その意図を予測することでランキングの質を向上できると報告した [7]。

また、EC サイト上の検索ではなく Web 検索におけるクエリやクリック情報を分析し、どのような意図で商品に関する検索がなされているのかを明らかにしようとする研究はいくつか行われている [3] [9] [13]。Guo らはスマートフォンのタッチやスワイプの情報を特徴量とし、ユーザの検索意図を分類するモデルを提案した [3]。Rao らは Bing に投入された商品に関する検索クエリを Comparison, Informational, Navigational, Support, Transactional に分類し、各検索意図での検索行動の違いを報告した [9]。Zhao らの研究ではユーザが過去に投入したクエリとクリックしたカテゴリをもとに、検索されたクエリがどの商品カテゴリに属するかをリアルタイムで予測するモデルを提案した [13]。

既存研究では Web 検索クエリを用いて詳細な検索意図への分類をしていなかった。特にユーザの商品購買後のクエリは、詳細な分類がなされていない。Web 検索クエリを用いて既存研究よりも詳細な分類することで、ユーザの検索行動について深く知ることができると考えられる。

### 2.2 购买前・購買後の分析

ユーザが商品を購入する前と購買した後で、クエリや Web 上での行動が変化するかを明らかにする研究がなされている。Kuchy らの研究では EC サイト上の製品に関する質問が购买前と購買後で変化するか分析している [6]。Zhang らは商品購買後のユーザの行動の違いを得るため Windows 10 と Xbox といった 2 つの商品を対象に購買前後での検索回数が増えるクエリについて分析している [12]。

既存研究では「购买前」、「購買後」という购买前か後かという点にだけ注目していたが、本研究では購買に至るまで、さらに購買した後にどのような変化をしていくのかを分析し、ユーザが求める情報の変化を明らかにする。

### 2.3 消費者購買プロセス

ユーザの消費者購買行動プロセスとして、AIDMA や AIS-CEAS, AISAS が提案されている。AIDMA では Attention (注意), Interest (関心), Desire (欲求), Memory (記憶), Action (行動) という 5 つの行動に分類している。

また、Web 検索ログを用いてユーザの購買における状態についての研究がなされている。Jansen らは検索キーワードを Awareness, Research, Decision, Purchase の 4 段階に分け、その意図による購買行動やクリックの違いを分析した [5]。

Web 検索における消費者購買行動について、既存の研究は少ない。本研究では、語の変化に着目することで Web 検索においてユーザの状態がどのように変化していくのかプロセスを分析する。

## 3 使用データ

本節では、本研究で使用したデータについて説明する。本研究では Yahoo!検索が提供する **Web 検索ログ**と Yahoo!ショッピングでのコンパクトデジタルカメラとデジタル一眼レフに関する商品の**購買データ**を用いる。

Yahoo!検索が提供する Web 検索において、2016 年 10 月 1 日から 2017 年 10 月 31 日の Web 検索ログを用いた。日常的に Web 検索を利用しているユーザを対象とするため、毎月 10 回以上 Web 検索を行っているユーザに分析対象を絞った。ユーザのクエリの変化に注目して分析するため、今回はページのクリックや滞在時間といった情報は用いない。

Yahoo!ショッピングにおいて、2016 年 10 月 1 日から 2017 年 10 月 31 日の購買データを用いた。本研究ではコンパクトデジタルカメラ、もしくは、デジタル一眼レフカメラに関する商品を対象期間中に購買したユーザについて分析する。カメラは比較的高価な商品であり、購買する際にユーザが比較や検討をしている様子がクエリに現れると考えられるため、本研究では

カメラを対象に分析する。また、Yahoo!ショッピングにおいて、2016年10月1日から2017年10月31日の期間にカメラを購入したユーザのことを本研究では**カメラ購買ユーザ**と呼ぶ。カメラを複数回購買したユーザについては、2016年10月1日から2017年10月31日の中で一番最初に購買した商品のデータを用いる。また、カメラ購買ユーザの分析期間としては、購買1週間前から購買1週間後のWeb検索クエリを対象に分析した。Web検索ログと購買データを結び付けることでユーザがどのように検索して商品の購買に至ったか、また商品購買後にユーザがどのような情報を求めているのかという分析が可能となる。

#### 4 検索回数とクエリに使用される語の時系列分析

はじめに、カメラの購買前後でユーザの検索回数は変化していくのか、また変化するのであればどのように変化していくのか分析する。次に、特定の時間区間にクエリに頻出する語は存在するのか、また頻出する語にはどのようなものがあるか分析する。

##### 4.1 購買1週間前から購買1週間後の検索回数の推移

本節では、カメラの購買前後において、Web検索での検索回数がどのように推移するのかを分析する。カメラ購買ユーザの購買1週間前から購買1週間後の検索回数の推移を図1に表す。図1では、購買1週間前から購買1週間後における24時間ごとの検索回数を2016年10月1日から2017年10月31日までの1日ごとの平均検索回数で割った値を表している。また、2016年10月1日から2017年10月31日までの1日ごとの平均検索回数を1とし、図中に赤点線で表している。

図1から、購買24時間前から購買直前にかけて、検索回数が一番多くなることが分かった。Windows10やXboxについて扱った既存研究でも同様に、購買直前に検索回数が増加することが示されている[12]。購買する商品を検討する上で、参考となる情報を求め、普段よりも検索するため、購買前に検索回数が増加することが考えられる。また、購買後においても購買直後から購買24時間後の検索回数は年平均検索回数よりも多いことが分かった。購買後にもユーザがカメラの購買に関して検索を行うことが考えられる。

##### 4.2 オッズ比に基づく語の分析

本節では特定の時間区間において、クエリに頻出する特徴的な語を分析する。具体的には購買1週間前から購買1週間後において、特定の時間区間によく出現する語をオッズ比を用いて抽出する。3節のデータセットから全カメラ購買ユーザにおける、購買1週間前から購買1週間後のすべてのクエリ集合を  $Q = \{q_1, q_2, \dots, q_n\}$  とする。本研究では、クエリは語の集合とみなす。購買1週間前から購買1週間後のすべてのクエリ集合  $Q$  に含まれる、クエリに出現するすべての語の集合を  $W$  とする。あるクエリ  $q$  における語  $w \in W$  の出現回数を  $\text{tf}(w, q)$  と表す。クエリ  $q$  において検索が行われた時間は、 $\text{time}(q)$  で表されるものとする。この時、特定の時間区間  $t$  におけるクエリ

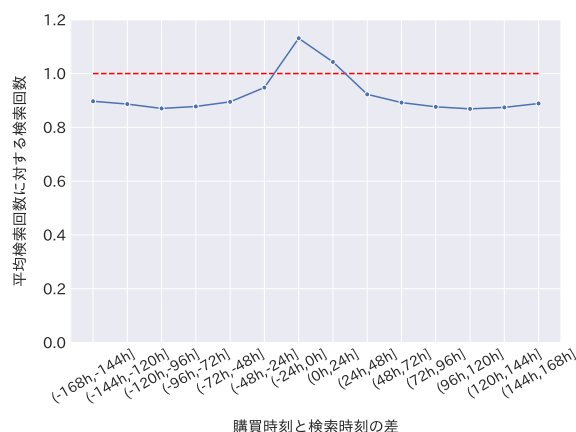


図1 購買1週間前から購買1週間後の24時間ごとの検索回数を2016年10月1日から2017年10月31日における1日ごとの平均検索回数で割った値。図中の赤点線は2016年10月1日から2017年10月31日における1日ごとの平均検索回数を表す。

集合  $Q_t$  を以下のように表す。

$$Q_t = \{q | q \in Q, \text{time}(q) \in t\}$$

時間区間  $t$  において、クエリ中に語  $w$  が含まれる出現頻度をクエリ集合  $Q_t$  を用いて、以下のように表す。

$$\text{tf}(w, Q_t) = \sum_{q \in Q_t} \text{tf}(w, q)$$

同様に、購買1週間前から購買1週間後における語  $w$  の出現頻度は  $\text{tf}(w, Q) = \sum_{q \in Q} \text{tf}(w, q)$  と表される。これらを用いて、購買1週間前から購買1週間後に検索された全単語  $W$  の総出現回数中に語  $w$  が出現する確率  $p(w)$  と、区間  $t$  において検索された語の総出現回数中に語  $w$  が出現する確率  $q(w, t)$  を求め、 $p(w)$  と  $q(w, t)$  を用いて、オッズ比を求める。

まず、購買1週間前から購買1週間後における語  $w$  の出現確率  $p(w)$  を以下の式で求める。

$$p(w) = \frac{\text{tf}(w, Q)}{\sum_{w' \in W} \text{tf}(w', Q)}$$

次に、区間  $t$  における語  $w$  の出現確率  $q(w)$  を以下の式で求める。

$$q(w, t) = \frac{\text{tf}(w, Q_t)}{\sum_{w' \in W} \text{tf}(w', Q_t)}$$

ここで、 $W$  はクエリ集合  $Q_t$  に含まれるすべての語集合である。 $p(w)$  と  $q(w, t)$  を用いてオッズ比  $\text{odds}(w, t)$  を以下の式で求める。

$$\text{odds}(w, t) = \frac{q(w, t)/(1 - q(w, t))}{p(w)/(1 - p(w))} \quad (1)$$

オッズ比  $\text{odds}(w, t)$  は購買1週間前から購買1週間後と比べ、区間  $t$  において語  $w$  が出現しやすいとき、また、検索ユーザ数や検索回数が少なく、区間  $t$  だけで語  $w$  が多く調べられているときに値が大きくなる。より多くのユーザがクエリに用いてい

表 1 特定の時間区間に頻出する語.

順位	(-72h,-48h]	(-48h,-24h]	(-24h,0h]	(0h,24h]	(24h,48h]	(48h,72h]
1	バリスタ	デジタル一眼レフ	190	互換バッテリー	お問い合わせ番号	マイクロ sd
2	取り込み	宇多田ヒカル	テレマルシェ	レンズプロテクター	佐川急便	gateway
3	蛍光灯	びっくりドンキー	コンパクトデジタルカメラ	レンズフィルター	伝票番号	elecom
4	栽培	ネットショップ	トイカメラ	液晶フィルム	news	pentax-da
5	女子高生	nicon	xp120	sdx	nfc	糸
6	cs	防塵	dsc-wx810	プロテクター	sagawa	枚数
7	似てる	手	tough	保護フィルター	クロネコ	nfc
8	なか卯	5のつく日	a10	付属品	追跡	観月あこ
9	有楽町	dmc-tz85	防水カメラ	レンズカバー	ヤマト運輸	sagawa
10	子役	x5	ixy	class10	佐川	bar

る語を対象とするため、本研究では 100 人以上がクエリに用いている語のみを分析の対象とした。

式 (1) で定義したオッズ比を用いて、購買 72 時間前から購買 72 時間後を 24 時間ごとに区切った際に各時間区間に頻出する語を求める。それぞれの時間区間において頻出する語をオッズ比を用いてランキングした結果を表 1 に表す。たとえば、表 1 において、(-24h,0h] という列の 1 位にある「190」という語は、購買 24 時間前から購買直前の間において、クエリ中に含まれる語の中でオッズ比が最も高く、時間区間の中で最も頻出する語が「190」であることを表している。

はじめに、商品購買前において頻出する語について注目する。購買 48 時間前から購買 24 時間前においては、「デジタル一眼レフカメラ」のようなカメラの種類についての語や、カメラの機能である「防塵」について、カメラの型番名である「dmc-tz85」が上位に現れた。また、購買 24 時間前から購買直前において頻出する語上位 10 件には、「xp120」や「dsc-wx810」、「a10」といったカメラの型番に関する語が頻出していることが分かった。

次に、商品購買後において頻出する語について見ていく。購買直後から購買 24 時間後に頻出する語には、「レンズプロテクター」や「液晶フィルム」といった補完品に関する語が頻出することが分かった。ユーザは購買直後に商品と同時に使用する商品の購買を検討していると考えられる。さらに、購買 24 時間後から購買 48 時間後には「伝票番号」や「佐川急便」といった配達に関する語が上位 10 件中 8 件に見られ、ユーザが商品の追跡や問い合わせを行っていることが考えられる。購買からの時間に応じて、クエリ中に頻出する語が異なることが分かった。

## 5 カメラの購買前後に特徴的な語の抽出と購買ラベルへの分類

本節では、カメラの購買前後のクエリ中に特徴的に出現する語を抽出する。さらに、それらの特徴的な語がどのような情報を求めるため利用されるのかという観点から分類する。本研究では、この分類を**購買ラベル**と定義する。

カメラを購買するユーザは、カメラ購買前にはカメラの価格を調べたり、カメラの人気を調べたりすることが考えられる。また、カメラ購買後には、ネットショッピングでの配達状況を

追跡しようとしたり、カメラで使用する SD カードについて調べたりすることが考えられる。本研究ではまず、このようなカメラ購買前後のクエリにおいて特徴的に利用される語を明らかにする。

### 5.1 分類器を用いた購買前・購買後に特徴的な語の抽出

本研究では、機械学習の分類器を用いることで、カメラの購買前、購買後のクエリに含まれる特徴的な語を抽出する。また、抽出した語にはどのような種類があるのか分類する。本節では、その分析手法の概要を説明する。

あるカメラ購買ユーザにおいて、購買前や購買後にどのような語がクエリに含まれているかについて Web 検索ログと購買データを組み合わせることによってわかる。購買前や購買後のクエリに使用される語の中からカメラの購買に特徴的な語を抽出することを考える。たとえば、クエリ中に「おすすめ」や「価格」という語が含まれていると購買前、「カメラケース」や「説明書」といった語が含まれていると購買後といったように判断できると考えられる語である。本研究ではある語がクエリ集合に含まれた時、クエリ集合が購買前か購買後かを判断できる語をカメラの購買前後のユーザの状態や目的を強く反映した語と考え、カメラの購買に特徴的な語とする。

カメラの購買前や購買後に特徴的な語を抽出するために、購買前に検索されたクエリ集合、購買後に検索されたクエリ集合から特徴ベクトルを作成し、入力した特徴ベクトルが購買前か購買後か分類する分類器を作成する。このような分類器において、分類に貢献する語を分析することで、クエリ集合が与えられた際に、購買前か購買後かを判断するため重要となる語がわかる。購買前か購買後かを判断するために重要となる語は、はじめに述べたカメラの購買に特徴的な語であると考えられる。そのために以下の 4 点に取り組む。また、次節ではこれらの手順について詳しく説明する。

(1) 各カメラ購買ユーザにおける、カメラの購買に関する購買前のクエリ集合、また購買後のクエリ集合を抽出する。

(2) 各カメラ購買ユーザにおける、カメラ購買前のクエリ集合、また購買後のクエリ集合を特徴ベクトル化する。

(3) クエリ集合から作成した特徴ベクトルに基づき、購買前のクエリ集合か購買後のクエリ集合かを判断する分類器を構

築する。

(4) 分類器から购买前か購買後の分類に貢献する語を調べ、その語をカメラの購買に特徴的な語とする。

5.2節では上記の手順に従い、カメラの購買に特徴的な語を抽出する。5.3節では、カメラの購買に特徴的な語にはどのような種類があるのか、人手で分類する。

## 5.2 カメラの購買前後で特徴的に利用される語の抽出

本節では前節で述べた手順に従い、カメラの購買に特徴的な語を抽出する。

まず、各カメラ購買ユーザにおける、カメラの購買に関する购买前、購買後のクエリ集合を抽出する方法について説明する。はじめに商品購買データや価格.com<sup>1</sup>から、カメラの商品名に関する語を抽出した。カメラの商品名に関する語として「nikon」や「canon」といったメーカー名、「eos」、「lumix」といったシリーズ名、「c200」や「ixy180」といった型番に該当する語を準備した。語は英語表記のみでなく、カタカナ表記やひらがな表記も用意した。

次に、カメラの購買に関する购买前のクエリ集合を抽出した。購買72時間前から購買直前に検索されたクエリのうち、カメラの商品名に関する語を含むクエリを抽出した。抽出したクエリ集合から、購買72時間前から購買直前において同一ユーザが30分以上間隔をあげずに投入したクエリ集合をセッションとして分割した[4]。分割したセッション中にカメラの商品名に関する語が含まれるクエリがあれば、そのセッションをすべて抽出した。あるユーザ  $u$  において、抽出したセッション中に含まれるクエリ集合を购买前のカメラの購買に関するクエリ集合  $Q_{before}^u$  とする。

また、购买前のカメラの購買に関するクエリ集合  $Q_{before}^u$  と同様に、購買直後から購買72時間後のクエリを対象にした、カメラの購買に関する購買後のクエリ集合  $Q_{after}^u$  を抽出した。

次に、得られたクエリ集合から分類器の入力となる特徴ベクトルを作成する。 $Q_{before}^u, Q_{after}^u$  中に含まれる語それぞれに対してTF-IDFを用いて重みづけを行い、 $Q_{before}^u, Q_{after}^u$  に対する特徴ベクトルを作成した。まず、TF値を求める。 $r \in \{before, after\}$  とするとき、クエリ集合  $Q_r^u$  中のクエリ  $q$  に含まれる語  $w$  の個数を  $tf(w, q)$  とする。クエリ集合  $Q_r^u$  に含まれる語  $w$  の個数  $tf(w, Q_r^u)$  は以下の式で求める。

$$tf(w, Q_r^u) = \sum_{q \in Q_r^u} tf(w, q)$$

次に、IDF値を求める。本研究で分析対象のユーザ集合を  $U$  とする。 $Q_{before}^u$ 、または、 $Q_{after}^u$  中のクエリ  $q$  に含まれる語  $w$  におけるIDF値  $idf(w)$  を以下の式で求める。

$$df(w) = |\{Q_r^u \mid r \in \{before, after\}, u \in U, tf(w, Q_r^u) \geq 1\}|$$

$$idf(w) = \log \frac{1 + N}{1 + df(w)} + 1$$

ここで、 $N$  は分類対象とするクエリ集合の個数であり、

$N = 2|U|$  である。TF値  $tf(w, Q_r^u)$ 、IDF値  $idf(w)$  を用いて、クエリ集合  $Q_r^u$  における語  $w$  のTF-IDF値  $tfidf(w, Q_r^u)$  は以下の式で求める。

$$tfidf(w, Q_r^u) = tf(w, Q_r^u) \cdot idf(w)$$

カメラを購買した7,531ユーザについて、TF-IDFの重みづけで作成したクエリ集合の特徴ベクトルを用いて、购买前を正例、購買後を負例とし、ロジスティック回帰により分類器を構築した。訓練データ:テストデータの割合を8:2とし検証を行った結果、正解率0.61、適合率0.52、再現率0.24、 $F_1$ 値0.33であった。構築した分類器から、回帰式において回帰係数の絶対値が大きい語を、係数が正の語、負の語それぞれ100語抽出した。回帰係数の絶対値が大きい語で係数が正の語はその語がクエリに含まれることで购买前と分類されやすくなる語、係数が負の語はその語がクエリに含まれることで購買後と分類されやすくなる語である。係数の値が正で高くなる語には店舗である「キタムラ」や商品の「価格」、係数の値が負で高くなる語には補完品である「ケース」や「sdカード」、配送会社の「佐川急便」といった語が抽出できた。抽出した語は、购买前や購買後に分類器を用いて抽出した語とカメラの商品名に関する語を合わせて、カメラの購買に関する語と定義する。

## 5.3 抽出した語の分類

本節では、どのような種類の情報を求めているのかという観点からカメラの購買に関する語を購買ラベルへ分類する。これにより、购买前、購買後の検索においてどのような種類の情報が検索されているのかを明らかにする。

検索行動とユーザの満足度の関係について調査した既存研究[11]では、特定の商品に絞り込んでいるかや商品の評価や比較、ブラウジングなどどのような目的で検索を行っているかという観点から、検索行動を分類した。本研究では、既存研究を参考に「カテゴリ」、「メーカー」、「シリーズ」、「型番」、「動機」、「購買」、「比較」、「評価」、「ブラウジング」、「付属品・補完品」、「アフターサービス」、「その他」といった12種類の購買ラベルを準備した。3名の評価者が5.2節で抽出した語集合に対し、それぞれ独立で購買ラベルの付与を行った。購買ラベルを付与できなかった語に対し、3名で追加の購買ラベルを用意した。「機能・特徴」、「配達」、「価格」、「購買サイト・店舗」、「付属品の情報」、「使用方法」、「その他(カメラに関すること)」といった7種類の購買ラベルを追加し、合計16種類の購買ラベルを用意した。カメラの購買に関する語に、再び3名の評価者がそれぞれ独立に合計16種類の購買ラベルを付与した。購買ラベルについて意見が分かれた際は多数決で決定し、多数決で決定できなかった場合は議論を行い決定した。

作成した購買ラベルを表2に示す。表2の「購買ラベル」の列には作成した購買ラベルを挙げている。「定義」の列には購買に関する語の中で、どのような語に購買ラベルを付与したのかを記している。「付与される語」の列に実際に分類された語の例を挙げている。付与した購買ラベルは合計16種類となった。

既存研究では主に购买前のユーザの検索行動に焦点を当てて

1: 価格.com: <https://kakaku.com/>

表 2 購買ラベルと定義, 付与する語の例.

購買ラベル	定義	付与される語の例
シリーズ	商品のシリーズ名に関する語.	coolpix, サイバーショット
カテゴリ	カメラの種類に関する語.	コンデジ, ミラーレス一眼
型番	商品の型番に関する語.	d5300, eos7d
メーカー	商品のメーカー名に関する語.	キヤノン, ニコンカメラ
機能・特徴	カメラ商品に関する機能や特徴に関する語.	ホワイト, 防水
購買サイト・店舗	商品を購入するためのサイト・店舗名に関する語.	楽天, ヤマダ電機
価格	商品の価格に関する語.	価格ドットコム, 安い
アフターサービス	商品の保証や購買後の行動における情報に関する語	保証, 製品登録
付属品の情報	付属品に関する機能や特徴, 情報に関する語.	12mm, サンワサプライ
評価	商品に対する評価・評判に関する語.	おすすめ, 人気
ブラウジング	商品の販売方法に対する情報に関する語.	割引, 5のつく日
付属品・補完品	カメラと一緒に使う商品に関する語.	充電器, ストロボ
配達	商品の配送に関する語.	佐川急便, 追跡
使用方法	商品の使用方法や使用に関する語.	応用撮影ガイド, 使い方
その他 (カメラ)	カメラに関する語だが, 上記以外に分類される語.	中古, 後継機
その他	カメラに関係しない語	

いた. 本研究で行った分類から, 「配達」や「使用方法」といった購買後のクエリに多く含まれると考えられる語における分類を明らかにすることができた. ユーザが購買前後において, 様々な情報を求めながら検索することが明らかとなった.

## 6 購買ラベルに基づくクエリの時系列分析

本節では 5.3 節で作成した購買ラベルに分類された語がいつクエリに用いられるのか分析する. 購買ラベルに分類された語がいつごろクエリに用いられるかを理解することで, 商品購買のタイミングや購買後にユーザが求める情報がわかる.

はじめに, 購買 1 週間前から購買 1 週間後の間を 24 時間ごとに分割し, 購買ラベルに分類された語が, クエリ中にいつごろ多く用いられるかを分析する. 次に購買に至るまでのクエリ, 購買後からのクエリにそれぞれ着目し, 購買ラベルに分類した語がクエリ中にいつごろ多く用いられるか分析する.

### 6.1 24 時間ごとの語の出現回数の推移

本節では購買 1 週間前から購買 1 週間後にかけて 24 時間ごとにおいて, 購買ラベルに分類された語がクエリ中にいつごろ多く用いられるかを分析する. このことを分析するため, 各購買ラベルにおいて, 購買 1 週間前から購買 1 週間後における出現割合を用いる. 出現割合は各購買ラベルに含まれる語の購買 1 週間前から購買 1 週間後における出現回数に対し, 特定の時間区間での購買ラベルに含まれる語の出現回数を割った値である. 図 2 に, 購買 1 週間前から購買 1 週間後における各購買ラベルの出現割合の推移を示す.

図 2 は, 購買ラベルが「シリーズ」において購買時刻と検索時刻の差が区間  $(-24h, 0h]$  に含まれるとき, 出現割合が 0.28 となっているのは, 購買 1 週間前から購買 1 週間後の「シリーズ」に分類される語の総出現回数を 1 とした時, 購買 24 時間前から購買直前における出現割合が 0.28 となることを表している.

まず, 購買 1 週間前から購買直前において, 購買ラベルに分

類された語がクエリ中に出現した回数の推移について述べる. 購買 24 時間前から購買直前において, 「価格」や「アフターサービス」, 「購買サイト・店舗」の出現割合が 0.1 以上あり, 期間の中で最も大きくなっている. さらに, 購買 24 時間前から購買直前における「シリーズ」, 「カテゴリ」, 「型番」, 「メーカー」の出現割合は 0.2 以上となっている. 「評価」と「ブラウジング」は購買後よりも購買前の方が出現割合が高くなっている.

次に購買直後から購買 1 週間後において, 購買ラベルに分類された語のクエリ中での出現回数の推移について述べる. 「付属品・補完品」については購買直後から購買 24 時間後にかけて出現割合が 0.16 となり最も高い出現割合となっている. さらに「配達」に関しては購買直後から徐々に検索されるようになり, 購買 48 時間後には出現割合が 0.20 となっていることが分かった. これは購買 48 時間後に商品が発送され, 商品の追跡などが行われているということが予想できる. また, 購買前と比較し, 購買後には「使用方法」の出現割合が高くなっていることが分かった. 商品が購買 72 時間後以降に商品が届き, 実際に使う際に検索をしていると考えられる.

### 6.2 購買前後の語の出現回数の推移

本節では購買前, 購買後のクエリに着目し, 購買ラベルに分類された語がクエリ中にいつごろ多く用いられるかを分析する. 6.1 節での分析では, 時間ごとの分析しか行うことができず, 個々のユーザの推移については注目できていない. 購買 1 週間前から毎日 Web 検索を用いて情報を収集するユーザや購買直前に 3 時間以内で検索を終わらせ, 商品を購入するユーザも考えられる. そこで, 購買前にカメラの購買に関する語を含む検索を 5 回, また, 購買後にカメラの購買に関する語を含む検索を 5 回行ったユーザを対象として, 購買ラベルの出現割合の推移を分析する.

はじめに購買 1 週間前から購買直前までの間, 購買直後から購買 1 週間後それぞれで 5.2 節で作成したカメラの購買に関する語を含むクエリをちょうど 5 回投入したユーザのクエリを抽



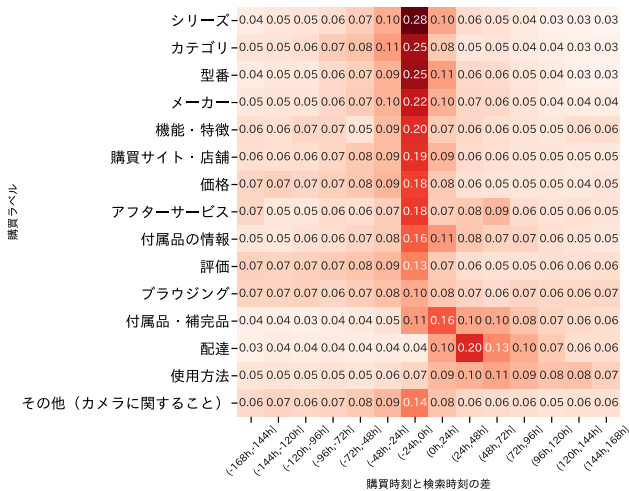


図 2 購買 1 週間前から購買 1 週間後、24 時間ごとにおいて、購買ラベルに分類された語がクエリ中に出現する頻度。図中の数値は購買ラベルごとに、商品購買 1 週間前から購買 1 週間後の総出現回数を各時間区間での出現回数で割った値である。

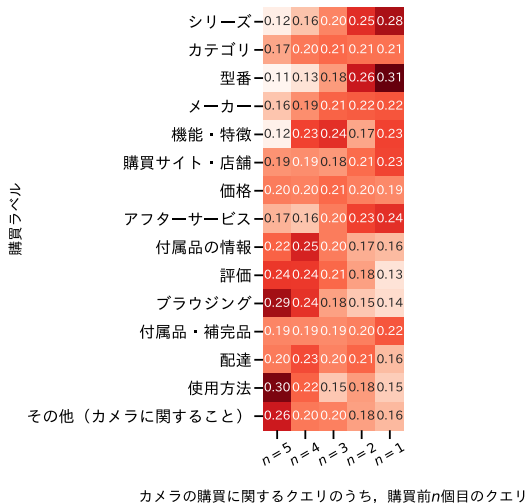


図 3 購買前にカメラの購買に関する検索を 5 回行ったユーザーにおける、購買ラベルごとの、クエリ中に出現する語の割合。図中の数値は、購買ラベルごとに、5 つのクエリ全体での出現回数に対する、対象のクエリでの出現回数を割った値である。

出する。購買ラベルごとにクエリ全体での総出現回数に対する、対象のクエリでの出現割合を購買前後それぞれにおいて求めた。図 3 に購買 1 週間前から購買直前に 5 回、図 4 では購買直後から購買 1 週間後に 5 回、カメラの購買に関する語を検索したユーザーに対する購買ラベルの出現割合の推移を示す。たとえば、図 3 において購買ラベルが「シリーズ」でカメラの購買に関するクエリのうち購買前  $n$  個目のクエリが  $n = 3$  のとき、値が 0.20 となっている。これは、購買前にカメラに関する語を 5 回検索したユーザーにおいて、購買ラベル「シリーズ」についての総出現回数を 1 とした時、購買から 3 個前のカメラに関する語を含むクエリの出現割合が 0.20 であるという意味である。

図 3 について、どのように購買ラベルが推移していくかを述

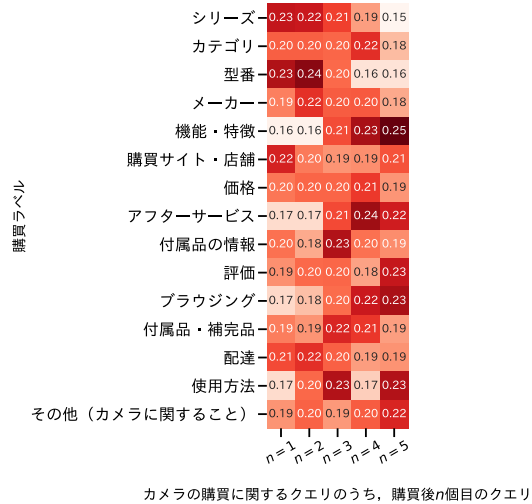


図 4 購買後にカメラの購買に関する検索を 5 回行ったユーザーにおける、購買ラベルごとの、クエリ中に出現する語の割合。図中の数値は、購買ラベルごとに、5 つのクエリ全体での出現回数に対する、対象のクエリでの出現回数を割った値である。

べる。「シリーズ」、「カテゴリ」、「型番」、「メーカー」についての出現割合が購買に近づくにつれて大きくなる。また、「人気」や「おすすめ」といった語が含まれる評価については、購買に近づくにつれて出現割合が小さくなっていくことが分かった。「通販」や「割引」といった語を含む「ブラウジング」や「使用方法」についての出現割合も小さくなっていくことが分かった。

図 4 での購買ラベルに分類された語の出現割合について述べる。購買後、「シリーズ」や「型番」についての出現割合が徐々に減少していくことが分かった。

## 7 議論

本研究では、Web 検索クエリとオンラインショッピングでの購買履歴を用いて、購買前と購買後のクエリにおける語の変化について、3 つのリサーチクエスチョンを明らかにすることに取り組んだ。

RQ1 について、購買前後における検索回数の推移とオッズ比を用いて特定の時間区間に頻出する語の分析を行った。Web 検索での検索回数は、購買 24 時間前から購買直前にかけて検索回数が最も多くなる。購買直前にユーザーが多様な情報を取得するため、クエリを大量に投入していることが考えられる。購買 72 時間前から購買 48 時間前には「栽培」や「バリスタ」のようにカメラの購買に関係すると考えられる語は少なかった。このことから購買 72 時間前から購買 48 時間前には、多くのユーザーはカメラに関する検索をまだしていないことが考えられる。また、購買 24 時間前から購買直前にかけて「xp120」や「dsc-wx810」といったカメラの商品名を含んだ検索が多くなることが分かった。購買 24 時間後から 48 時間後には「佐川急便」や「伝票番号」、「ヤマト運輸」などの語が上位となっていた。

RQ2 について、カメラの購買に関する語を機械学習を用いて抽出し、抽出した語がどのような種類に分かれるかを分類し

た。抽出した語には「メーカー名」や「シリーズ名」といった商品に関する語以外に、「使用方法」や「配達」に関する語も現れた。ユーザが購買前後に様々な情報を求めていることが明らかとなった。

RQ3 では RQ2 で分類した語の種類を購買ラベルと定義し、購買ラベルごとの出現割合の推移を分析した。「メーカー名」や「シリーズ名」、「型番」などが購買 24 時間前から購買直前に大きくなるのが分かった。購買後においては、特徴的な購買ラベルの推移を見ることができた。購買直後から購買 24 時間後には「sd カード」や「レンズプロテクター」といった語を含む「付属品・補完品」、購買 24 時間後から購買 48 時間後には「佐川急便」や「追跡」といった語を含む「配達」、購買 48 時間後から購買 72 時間後には「取説」や「使い方」といった語を含む「使用方法」の検索が多くなっていくことが分かった。このような詳細な動きを理解することにより、購買後にユーザが求める情報を予測でき、適切な広告配信が期待できる。また、購買直前や購買直後のクエリにおける購買ラベルの出現割合の推移を分析した。購買直前には、「人気」や「おすすめ」を含む「評価」の出現割合が小さくなっていくことが分かった。これは、ユーザが始めに、「人気」や「おすすめ」をクエリに含むことで商品を比較し、得られた情報をもとに購買する商品を絞り込んでいるからだと考えられる。購買直後には「メーカー」や「シリーズ」、「型番」についての出現割合が徐々に小さくなっていくことが分かった。このような知見を活かすことでユーザがいつごろ商品を購入するかといった予測につながると考えられる。

次に、本研究での限界点について述べる。本研究ではユーザの過去の経験などを考慮した分析を行っていない。ユーザが分析対象期間の前に、既に購買しているどうかは考慮せずに、対象期間の中で最初に購買した時間を用いて分析を行った。過去にカメラを購入したことがあるユーザは、過去に購買した際の事前知識があるため、Web 検索によって求める情報が購買したことのないユーザと異なると考えられる。そのため、クエリに投入される語の種類や投入される時期も変化していくと考えられる。購買経験の有無に考慮した分析することができれば、さらに詳細なユーザの購買行動を知ることができたと考えられる。

本研究では作成した購買ラベルがカメラに関する検索をどの程度網羅できているかに関する分析を行っていない。カメラ購買ユーザのクエリを無作為に抽出し、本研究で作成した購買ラベルに分類することで Web 検索クエリにおける購買ラベルの網羅性を確認することができると考えられる。

## 8 まとめと今後の課題

本研究では EC サイトでカメラを購入したユーザを対象に、Web 検索クエリ中の語が購買前から購買後にかけて時間とともにどのように変化するかを調査した。はじめに、特定の時間区間において頻出する語はあるのかをオッズ比を用いて、分析し、カメラの型番が購買直前に現れることを明らかにした。次に、カメラの購買前後に特徴的に現れる語を分類し、「配達」や「付属品・補完品」といった語をクエリに用いることが明らかと

なった。また、分類した語が購買前後のいつごろクエリに用いられるのかを分析した。購買直後から購買 24 時間後には「付属品・補完品」、購買 24 時間後から購買 48 時間後においては「配達」、購買 48 時間後から購買 72 時間後においては「使用方法」についての検索が増加することが分かった。

本研究の知見を活かすことにより、ユーザは購買後に補完品や使用方法などについて検索を行っていることから、ユーザが求める情報を予測し、広告配信も可能になると考えられる。

今後は日用品やファッション、家具といった、カメラ以外のドメインの商品に対しても購買前と購買後の Web 検索クエリの変化を分析し、本研究での知見が異なるドメインにも対応できるのか検証していく。

**謝辞** 本研究は JSPS 科学研究費助成事業 JP18H03494, JP21H03774, JP21H03775, 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] A. Ashkan and C. L. Clarke. Characterizing commercial intent. In *Proc. of CIKM*, pp. 67–76, 2009.
- [2] A. Boteanu, E. Dutilleul, A. Kiezun, and S. Artzi. Subjective search intent predictions using customer reviews. In *Proc. of CHIIR*, pp. 303–307, 2020.
- [3] L. Guo, L. Hua, R. Jia, B. Zhao, X. Wang, and B. Cui. Buying or browsing?: Predicting real-time purchasing intent using attention-based deep network with multiple behavior. In *Proc. of KDD*, pp. 1984–1992, 2019.
- [4] J. Huang and E. N. Efthimiadis. Analyzing and evaluating query reformulation strategies in web search logs. In *Proc. of CIKM*, pp. 77–86, 2009.
- [5] B. Jansen and S. Simone. Bidding on the buying funnel for sponsored search and keyword advertising. *Journal of Electronic Commerce Research*, 12(1):1, 2011.
- [6] L. Kuchy, D. Carmel, T. Huet, and E. Kravi. “Did you buy it already?”, detecting users purchase-state from their product-related questions. In *Proc. of SIGIR*, pp. 1249–1258, 2021.
- [7] S. Manchanda, M. Sharma, and G. Karypis. Intent term selection and refinement in e-commerce queries. *arXiv preprint arXiv:1908.08564*, 2019.
- [8] G. Murga. Amazon Takes 49 Percent of Consumers’ First Product Search, But Search Engines Rebound. <https://blog.surveya.com/amazon-takes-49-percent-of-consumers-first-product-search-but-search-engines-rebound>.
- [9] N. Rao, C. Bansal, S. Mukherjee, and C. Maddila. Product insights: Analyzing product intents in web search. In *Proc. of CIKM*, pp. 2189–2192, 2020.
- [10] P. Sondhi, M. Sharma, P. Kolari, and C. Zhai. A taxonomy of queries for e-commerce search. In *Proc. of SIGIR*, pp. 1245–1248, 2018.
- [11] N. Su, J. He, Y. Liu, M. Zhang, and S. Ma. User intent, behaviour, and perceived satisfaction in product search. In *Proc. of WSDM*, pp. 547–555, 2018.
- [12] Q. Zhang, S. Hill, and D. Rothschild. Post purchase search engine marketing. In *Proc. of TheWebConf*, pp. 663–670, 2018.
- [13] J. Zhao, H. Chen, and D. Yin. A dynamic product-aware learning model for e-commerce query intent understanding. In *Proc. of CIKM*, pp. 1843–1852, 2019.
- [14] 中田祐誠, 村本直樹, 山本岳洋, 藤田澄男, 大島裕明. ウェブ検索ログからのカメラのオンライン購買行動予測. 人工知能学会論文誌, 36(1, W12-C.1):1–10, 2021.