

ネットショッピングにおける商品選択条件の推定

Estimating method on choosing the item in E-Commerce

江崎 健司[♥] 倉島 健[♦]
市川 祐介[♦] 高屋 典子[▲]
内山 匡

Kenji ESAKI Takeshi KURASHIMA
Yusuke ICHIKAWA Noriko TAKAYA
Tadasu UCHIYAMA

EC サイトにおいて自顧客のニーズを把握しマーケティングを行うことは重要である。本研究では EC の商品閲覧行動モデルを提案し、閲覧した商品の商品選択条件語(ISRW)を推定することを考える。提案する閲覧行動モデルは Focusing-IBB, Preference-IBB の二つに分類する。本研究は二つの特徴があり、閲覧行動は個人ごとに特徴が異なるため個人ごとに ISRW を推定することと、同一の商品であっても目的が異なれば ISRW が異なるため閲覧行動ごとに ISRW を推定する。

Recently, understanding of customer behavior, view or buy items etc., and their reasons are more important in E-commerce marketing. This paper proposes the customer item browsing behavior model (IBB) for estimating item selected reason words (ISRW). Our model captures two types of user behaviors, Focusing-IBB and Preference-IBB. This paper has two points of view. One is estimating ISRW for each user, the other is for session, because IBB is different from user and purpose.

1. はじめに

EC サイトにおける重要な課題は、自顧客のニーズ把握しマーケティング戦略に活用することである。これまでPOS データを利用し、ブランドのロイヤリティ推定や価格の受領範囲等、顧客の商品購買行動をテーマとした様々な研究が行われている。またECサイトでは商品閲覧行動の獲得が可能なため、個人の商品の検索から、閲覧、購買に至るまでの一連のプロセスをとらえることができるメリットがある。このメリ

ットによって実店舗で行うことが難しかった未購買の商品閲覧行動の分析が可能となる。商品閲覧行動を分析することで潜在的な商品購買行動を把握し、買おうか迷った商品、比較検討した商品の分析や、比較し購入したのか数日間迷って購入したのかなどの閲覧から購買に至るプロセスの分析も可能になり、顧客をより深く理解することができる。そこで本研究では、自顧客のニーズ把握やマーケティング戦略への活用のために、商品閲覧行動で閲覧した商品が選択された理由を商品選択条件語 (ISRW (Item Selected Reason Word)) として推定する。ISRWを推定する事で自顧客や店舗でよく選択条件となる語の可視化が可能になり、自顧客や店舗で推定した ISRW と一致する商品に注力するなどという戦略をとることができる。本研究は二つの特徴がある。一つ目は、同一の商品であっても選択した理由は個人ごとに異なるものと考え、自然であるため、個人ごとに ISRW を推定する。具体的には、商品閲覧行動モデルを個人ごとに適用する。二つ目は、商品選択の条件となる語は、商品の形状や素材、装飾特徴、感性表現など、ある商品を表現するWebページ上の記述すべてが対象となりうるが、商品閲覧行動 (Item Browsing Behavior; IBB) を考慮して ISRW を推定する。具体的には、目的にあう商品を具体的に絞りこんで閲覧する行動である Focusing-IBB, 目的にあう商品がはっきりせず閲覧する行動である Preference-IBB の二つに着眼する。本研究では商品閲覧の際の目的がない行動、例えば window-shopping などとは対象外とする。

本論文の構成は次の通りである。2章で従来研究の概要と提案手法との差異を述べる。3章で提案手法の概要と閲覧行動モデルの提案について述べる。4章では、実験で用いる評価データと提案手法の実装について述べる。5章で実験結果と考察について述べ、6章でまとめと今後の予定について述べる。

2. 関連研究

本研究の関連研究として、Web 閲覧履歴からの情報抽出研究とユーザ行動モデリング研究について述べる。

枝らは、Web 閲覧履歴におけるキーワード遷移に着目した Web ページ推薦システムを提案している[6]。次に閲覧するページの予測のために、過去に閲覧したページから抽出した重要語間の時系列パターンを抽出する研究であり、閲覧者が商品を選択した理由 (ISRW) を抽出することを目的とした本研究とは異なる。

Web 閲覧や購買履歴に基づき、PLSA (Probabilistic Latent Semantic Analysis) [7] や LDA (Latent Dirichlet Allocation) [8] などの潜在トピックモデルを用いてユーザの興味をモデル化する研究が存在する。Ahmed らは、短期的、長期的な興味を別々に扱うモデルを提案しており、全体のトレンドに加え、一日、一週間、一ヶ月と、異なる期間での潜在変数の分布の重み付き線形和でユーザの興味を表現している[1]。しかし、PLSA, LDA や Ahmed らの手法など、潜在トピックモデルに基づく解析においては、得られた潜在トピック、潜在変数を人間が解釈するのが困難であるという問題がある。それに対して、提案手法は、商品が選択された理由を言語 (商品選択条件語) として抽出できるメリットがある。

[♥] 正会員日本電信電話株式会社 NTT サイバーソリューション研究所 esaki.kenji@lab.ntt.co.jp

[♦] 正会員日本電信電話株式会社 NTT サイバーソリューション研究所 kurashima.takeshi@lab.ntt.co.jp

[♦] 非会員日本電信電話株式会社 NTT サイバーソリューション研究所 ichikawa.yusuke@lab.ntt.co.jp

[▲] 非会員日本電信電話株式会社 NTT サイバーソリューション研究所 takaya.noriko@lab.ntt.co.jp

非会員日本電信電話株式会社 NTT サイバーソリューション研究所 uchiyaama.tadasu@lab.ntt.co.jp

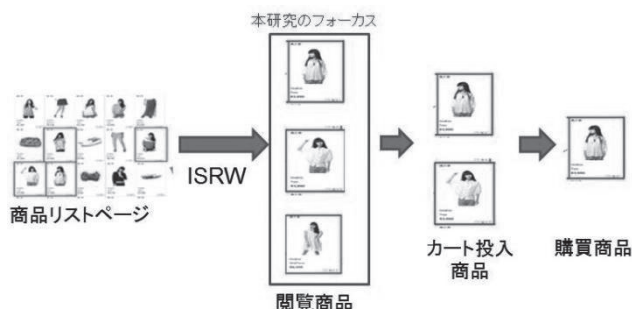


図1 商品閲覧ログにおける本研究のフォーカス

表1 商品閲覧行動の分類

Focusing-IBB	目的があり、 絞り込んだ条件がある	本研究の対象
Preference-IBB	目的があるが、 はっきりした条件はない	本研究の対象
その他	目的がない (Window-shopping 等)	本研究の対象外

3. 商品閲覧行動モデルを用いた個人ごとの閲覧商品の ISRW 推定法の提案

本章では商品閲覧行動モデルを個人ごとに適用し ISRW の推定を行う。本研究で用いる EC サイトの閲覧、購買に関する行動ログは約 800 名から収集し、その収集期間は約 1 年間（2010 年 11 月～2011 年 11 月）、収集した対象コマース数は 150 サイトである。本研究では商品閲覧に関する行動ログデータを用いて ISRW を推定する。本研究においては特に、商品詳細ページにフォーカスして ISRW の推定を行った（図 1）。商品詳細ページとは、サイト内のナビゲーション、検索等により表示された商品リストページから、顧客が選択し閲覧したページであり、商品の詳細に記述しており、かつ、カートに入れるなど購買の次プロセスに進むことのできるページを指す。

3.1 商品閲覧行動モデル

表 1 のように閲覧行動を 3 種に分類し、その中でも閲覧した商品に目的がある Focusing-IBB と Preference-IBB を対象とする。目的がない閲覧行動、例えば window-shopping などは本研究の対象外とする。目的がない場合は明確な理由が存在しない偶発的な場合であると考えられ ISRW の推定には有用とはいえないため対象外とした。

3.2 Focusing-IBB モデル

Focusing-IBB は目的にあった条件を具体的に絞り込んでいく閲覧行動であり、「豹柄」といった決まった ISRW を含む商品を連続して閲覧すると仮定しモデル化する。連続して閲覧されたことを同一セッションで閲覧されたと定義しスコアを算出する。有効なスコアを算出するために同一セッションの定めた商品以上で出現している商品特徴語を対象とする。ここでいうセッションとは、時系列的に連続してみられた Web ページの集合のことを指す[3]。ここで商品特徴語は予め用意しておいた商品特徴語辞書に含まれる語とする。商品特徴語辞書は、複数の EC サイト、ファッション情報サイト、雑誌記事等から係り受け関係をもとに商品を修飾している語、感性表現に相当する語を集めたものである。同一セッションに 3 商品以上で出現している商品特徴語を対象とした。Focusing-IBB に基づくスコア算出式を以下に示す。

表2 記号一覧

Focusing-IBB	
$N(i, u)$	同一セッション中の過去閲覧商品
K	最大カテゴリ距離
$Dis_{cate}(i, i')$	商品間のカテゴリ距離
$tf(w, i)$	商品 i での単語 w の出現頻度
Preference-IBB	
$\lambda(w, u)$	頻度パラメータ
$\mu(w, u)$	離脱パラメータ
$\Delta T(w, u)$	w を含む直前商品からの経過時刻
結合モデル	
a	重み付線形和パラメータ



図2 Focusing-IBB の概念図

$$S_N(w, i, u) = \sum_{i'=0}^{N(i, u)} \log \frac{K}{Dis_{cate}(i, i') + 1} * tf(w, i') \quad (1)$$

ここで、 $S_N(w, i, u)$ はあるユーザ u ごとに算出される商品特徴語 w の商品 i に対しての Focusing-IBB モデルにおけるスコアである。 K は最大カテゴリ距離、 $Dis_{cate}(i', i)$ は商品 i と i' のカテゴリ距離、 $tf(w, i')$ は商品 i' での商品特徴語 w の出現回数、 $N(i, u)$ は商品 i より以前の同一セッションの商品である。ここで、カテゴリ距離とは類似するカテゴリほど小さい値とするものとする。表 2 に本論文で用いる記号をまとめて示す。

Focusing-IBB モデルの概念図を図 2 に示す。今、赤い枠で囲んである商品について Focusing-IBB を算出する場合を考える。Focusing-IBB の場合、連続して閲覧された商品は同一の ISRW であると考えられるため、その商品より前に閲覧した同一セッションの商品を用いて ISRW を推定することができる。ただし、カテゴリが異なる場合は ISRW が異なる可能性があり、異なるカテゴリを閲覧する場合もある。図 2 の場合、同一セッション内でファー付が ISRW である商品が多く選ばれているため推定が可能である。図 2 では「シューズ」についての ISRW に、「バック」についての ISRW よりも「シューズ」の ISRW のほうが同一である可能性が高いと考え、カテゴリ距離に反比例する重みを用いてスコアを算出している。

3.3 Preference-IBB モデル

Preference-IBB は目的がはっきりしていない場合の閲覧行動であり、目的にあった条件で具体的に商品を絞り込むのではなく、「かわいい」という条件にあう「リボン」や「レース」などといった複数の ISRW を含む商品を閲覧する。Focusing-IBB は一時的な目的である場合でも推定する必要があるが、Preference-IBB モデルでは、個人の嗜好と一致する ISRW を含む商品の閲覧は、長期間にわたって頻出し継続していると仮定し、次式でスコアを算出する。有効なスコアを算出するために閲覧者が過去閲覧した定めた商品以上

に出現している商品特徴語を対象とする。本研究では 40 商品以上に出現している商品特徴語を対象とした。

$$S_P(w, i, u) = \frac{\lambda(w, u)}{\mu(w, u)} * e^{\mu(w, u)\Delta T(i)} \quad (2)$$

ここで、 $S_P(w, i, u)$ はあるユーザ u ごとに算出される商品特徴語 w の商品 i に対して Preference-IBB のスコアである。 $\lambda(w, u)$ は商品特徴語が出現する頻度の期待値、 $\mu(w, u)$ は離脱率、 $\Delta T(i)$ は商品 i とそれ以前に商品特徴語を含んだ商品との時間間隔である。図 2 では、“リボン”が ISRW である商品は同一セッションにはあまり出現していないが、過去のセッションでは頻出している。このように、明確な ISRW に加えて自分の好みの嗜好にあったものを選択する場合が Preference-IBB による選択になる。

式(2)はマーケティング分野で用いられている生存率モデルに基づき、商品特徴語が商品 i を閲覧したときに商品特徴語 w が出現する期待値をスコアとしている[5]。生存率モデルの仮定を ISRW となる商品特徴語の推定に適用するために次の 3 つの仮定を置いている。まず、商品特徴語が ISRW である商品の閲覧は過去商品特徴語を含む商品を閲覧した時刻に関係なくランダムに発生し、このランダムに発生するモデルはポアソン分布に従うと仮定する。次に、商品特徴語が ISRW となくなり離脱するかどうかは、過去商品特徴語が ISRW として生存していた時間に関係なくランダムに離脱し、このランダムに離脱するモデルは指数分布に従うと仮定する。最後に、ポアソン分布と指数分布のパラメータ $\lambda(w, u)$ 、 $\mu(w, u)$ は商品特徴語ごとに異なると仮定する。

3.4 結合モデル

商品に対する ISRW について、Focusing-IBB モデルと Preference-IBB モデルそれぞれのスコアを利用しそれらを結合したスコアを次式で算出する。

$$S(w, i, u) = a * S_N(w, i, u) + (1 - a) * S_P(w, i, u) \quad (3)$$

$S(w, i, u)$ は Focusing-IBB モデルと Preference-IBB モデルのスコアを重み付線形和したスコアである。次章で上記スコアが上位のものを商品 i の ISRW として推定したものを評価する。 a はモデル結合の重み係数である。本研究では $a=0.7$ とした。モデルの結合重みを変更する事によってどのような ISRW を推定するのかを調整することが可能である。

3.5 精度評価のための比較手法

Focusing-IBB モデルに対しては同一セッション中での出現頻度のスコアと RAKE[2]を比較手法として用いる。また、Preference-IBB モデルに対しては過去に閲覧した商品に語が出現した頻度、現在から過去に商品特徴語を含む閲覧した商品した時間によって減衰項を用いた出現頻度の総和を比較手法として用いる。結合モデルに対しては TF・IDF と RAKE を比較手法として用いる。

3.5.1 RAKE

RAKE は、キーワードを抽出するために提案された手法であり、文書集合からキーワード抽出をするのではなく一つの文書からキーワードを抽出できる特徴がある。本研究では一つの商品の商品詳細ページに記述があるテキストが短いため、同一セッションを一つの文書として扱い、次の式でスコアを算出する。

$$S_{RAKE}(w, i, u) = \frac{deg(w, i, u)}{freq(w, i, u)} \quad (4)$$

$freq(w, i, u)$ はパネル u が閲覧した商品 i を含むセッション中の商品群での単語 w の出現頻度、 $deg(w, i, u)$ は複合語中の出現頻度も合わせた出現頻度スコアである。リボンボーダーワンピースは、“リボン”、“ボーダー”、“ワンピース”と 3 単語の複合語になっているため頻度スコア 3 を加算する。複数の単語を用いた具体的で長いキーワードの方が重要という知見による。

3.5.2 減衰項を用いた出現頻度総和

出現頻度に時間減衰項を用いたスコア算出法を以下に示す[4]。

$$S_{Atten}(i, w, u) = \sum_{i'=0}^{N(i, u)} exp(-\tau \Delta T(i, i')) \quad (5)$$

ここで、 $N(i, u)$ は商品 i より以前の同一セッションの商品、 τ は減衰パラメータ、 $\Delta T(i, i')$ は商品 i とそれ以前に商品特徴語を含んだ商品 i' との時間間隔である。

3.5.3 TF・IDF

語の重要性を決める基本的な手法として TF・IDF があり、次式でスコアを算出する[4]。

$$S_{TF \cdot IDF}(i, w, u) = \log(1 + f(i, w, u)) * \log \frac{N}{df(w, u)} \quad (6)$$

$f(i, w, u)$ はパネル u が閲覧した商品 i を含むセッション中の商品に出現した単語 w の出現頻度、 df はパネル u が閲覧した単語 w を含む商品数である。 N は全商品数である。

4. 実験データと提案システム

4.1 実験で用いる評価データ

ISRW 推定結果を評価するため、3 章で述べた閲覧購買行動グのうち 2011 年 8 月からのログデータにおいて、同一セッションから 3 商品以上の商品詳細ページを閲覧し、かつ、商品詳細ページを含むセッションを 40 以上保持するモニタを候補者として選定した。閲覧していた本人に、対象商品ページを閲覧した際にその商品を選択した条件を回答してもらうアンケートを実施した。各モニタには最大 5 セッション、各セッションで最大 5 商品に対して選択肢を提示し、自身の選択条件に合致しているものにチェックを入れる方法をとった。該当する選択条件がない場合には、自由記述によって回答してもらった。提示した選択肢は、ブランド名と価格を含めて最大 20 語とした。提示した 20 語は、閲覧した商品詳細ページから抽出した商品特徴語と、そこから拡張した商品特徴語とした。語の拡張はテキストから抽出した商品特徴語と全商品詳細ページの記述から抽出した商品特徴語の共起頻度によって行った。図 3 にアンケート対象のセッションとそのセッションに含まれた商品と選択肢の例を示す。

では、1セッション目についての質問です。(全5セッション)

下の図はor1507様が08月23日にECサイトの、商品詳細ページへアクセスして閲覧された商品画像に並べたものです。各画像の下には、商品名・ブランド・サイズを表示しています。

その商品詳細ページにアクセスした理由として、当てはまるキーワード全てにチェックを入れて下さい。リストにないキーワードがあれば自由記述欄に記入して下さい。

※選択数は閲覧画面に記載の文章から、自動的にピックアップしたキーワードを表示しています。そのため、アイテムに関連しないものが表示される場合や、十分な数値が表示されない場合があります。アクセスした理由として適切なキーワードが数値にない場合は、自由記述欄にキーワードを記入して下さい。



ALL MY OWN
アクアガール/aqua
i/m256

一つ目(2011-08-23 21:45:34 閲覧)

☒ ディテール	☐ ボリューム	☐ ボウタイ
☒ フェミニン	☐ ライン	☐ キャザー
☐ マストアイテム	☐ ピンク	☐ ライトピンク
☐ フリル	☐ 縞	☐ レーヨン
☐ コットン	☐ 表地	☐ ブルー
☒ シフォン	☐ パープル	☐ 千鳥き
☒ アクアガール	☐ #4620	

/aqua
上記以外の理由(キーワードを自由記述)

☐ この商品には見覚えがない

図3 アンケート画面

このような評価実験でアンケート調査に回答したモニタは183名で、提案手法により推定した3502商品についてISRWとして合致しているかの評価データを得た。アンケートの結果、183名に提示した3502商品中1813商品が回答され5割弱の商品が提示した選択肢に回答した。ISRWが自由記述欄のみの場合は399商品であった。

商品閲覧行動モデルで用いたFocusing-IBBとPreference-IBBの仮定にあてはまった回答のうち提示したISRWにチェックをいれたのは、Focusing-IBBが158名、388商品、1666語、Preference-IBBが131名、191商品、659語であった。なお、商品が異なる場合は同一の語であっても別のものとして数えた。

4.2 評価

本研究では、精度と網羅性の二つの指標で手法を評価する。推定したISRWの精度を次式で評価し、網羅性をアンケートで回答があったISRWが推定可能な商品のうち何%の商品に対してISRWを推定できるかで評価する。

$$\text{精度} = \sum_s \frac{\sum_i \text{relevant}(i)_s * P@i_s}{N} \quad (7)$$

ここで、sはアンケートを行った商品、iは結合モデルで推定されたスコアが上位i番目のISRW、relevant(i)はi番目のISRWが正解であれば1をそうでないならば0を取る関数、P@i_sは商品sのi番目の商品選択条件のPrecisionである。

5. 実験結果と考察

5.1 評価データの傾向

アンケートによって得られた回答の傾向を分析する。まず、回答された語を分類したものを表3に示す。このうち選択肢によって回答されたもののなかでブランド名、価格が約半数を占めることがわかる。また、同一のISRWによって閲覧さ

表3 実験の回答結果傾向

名称	語数
選択肢による回答	4404語
ブランド名	1658語
価格	850語
同セッション3回以上出現	992語
同セッション3回未満出現	1894語
自由記述による回答	1432語
商品特徴語で記述可能な回答	1297語
商品特徴語で記述不可能な回答	145語

表4 実験結果

手法	精度	網羅性
Focusing: 出現頻度	0.42	21%
Focusing: RAKE	0.40	21%
Focusing: 提案手法	0.69	21%
Preference: 出現頻度	0.36	11%
Preference: 時間減衰	0.45	11%
Preference: 提案手法	0.49	11%
結合モデル: TF・IDF	0.48	31%
結合モデル: RAKE	0.38	31%
結合モデル: 提案手法	0.68	31%

れた場合、同一セッション上に共通の語を含むと仮定し、共通の語を含む商品数が3回以上かどうかで同一のISRWを区別すると約1:2となる。

5.2 実験結果

表4にアンケートで回答された語の推定精度を評価したものを示す。Focusing-IBBについて、同一セッション中での出現頻度をスコア(表4ではFocusing:出現頻度と記載)、RAKE、提案手法の精度は0.42, 0.40, 0.69だった。Preference-IBBについて、過去に閲覧した商品に語が出現した頻度(表4ではPreference:出現頻度と記載)、時間によって減衰項を用いた出現頻度の総和(表4ではPreference:時間減衰と記載)、提案手法の精度は0.36, 0.45, 0.49となった。そして結合モデルについて、TF・IDF, RAKE, 提案手法の精度は0.48, 0.38, 0.68となった。表4に各手法において正解データを含む商品のうち何%のISRWを推定できているかを評価したものを示す。Focusing-IBBについては21%、Preference-IBBについては11%、結合モデルについては31%となった。提案手法は、全商品のうち31%の商品について、0.68の精度でISRWを推定することが可能である。

5.3 考察

5.3.1 提案手法の有効性

Focusing-IBB, Preference-IBB, 結合モデルの全てにおいて提案手法が有効であった。Focusing-IBBで提案手法が有効だったのは特に今回取り扱ったファッション分野では、複合語など具体的なISRWよりもテイストを表すかわい、

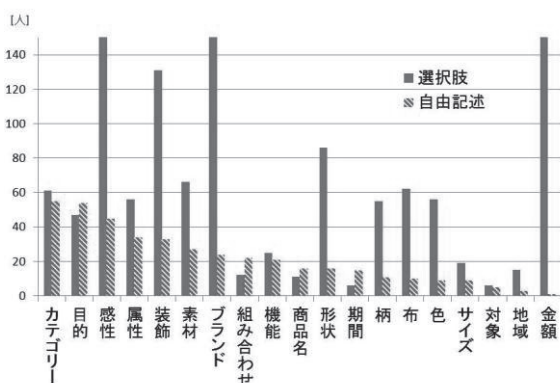


図4 ISRWの属性分析

リボンなど短い語であっても重要な語やブランドなどの固有名詞が回答となるケースが多かったことが原因だと考えられる。さらに、カテゴリごとに ISRW を推定する方法が有効であったため、精度改善につながったと考えられる。Preference-IBB に関する実験においては、提案手法も含め、時間減衰を考慮した手法のほうが、精度が高かった。提案手法においては、商品特徴語を含む商品が過去には閲覧しているが最近閲覧していない場合、好みや嗜好はすでに失われているとみなす。過去に多く閲覧しているが最近閲覧していない商品よりも、最近閲覧した商品に関する商品特徴語の方が個人の好みや嗜好を反映していると考えられる。このような仮定が正しかったため、精度向上につながったと考えられる。最後に、結合モデルは、Focusing-IBB と Preference-IBB という閲覧行動に関する複数の側面を考慮することができるため、比較手法に対する大きな精度向上につながったと考える。

5.3.2 ISRWの属性分析

図4にアンケート回答の属性の例を選択肢、自由記述に分けて示す。表5に属性の例を示す。選択肢のブランドと価格は描画の都合上で上限になっているが、属性がブランド名に関するものは766語、価格に関するものは386語、感性に関するものは209語である。ここで回答への属性の付与は人手で行った。回答された属性はさまざまで、選択時と自由記述の回答によって傾向が違ってくる。自由記述による回答の中で、商品特徴語で記述不可能な回答は1割程度であった。このことから、商品特徴語で記述できるものを対象とすることで十分であると考えられる。ここで、商品特徴語で記述不可能かどうかは人手でチェックし判定した。商品特徴語で記述不可能な例としては、凝ったデザイン、自分の定番などがあつた。

5.3.3 ISRWと提案の有効性

ISRWとして回答された語で最も多かったのはブランド名に関するものであるが、一回もブランド名を ISRW として回答しなかったパネルは115人存在した。また必ず ISRW にブランド名を回答したパネルは存在しなかった。そのため、ブランド以外で ISRW を推定することは意義があると考えられる。また、ISRW として2つ以下の属性のみを用いたパネルが102人、8つ以上の属性を用いたパネルが27人いた。このようにパネルに ISRW の個人差が存在するため個人ごとに ISRW を推定する事は重要であると考えられる。図5に Focusing-IBB と Preference-IBB それぞれで推定された属

表5 属性分析の例

属性	属性例
カテゴリー	トート、ヒール、財布、ピアス
目的	リゾート、トレッキング、山ガール、お買い得
感性	ムード、おしゃれ、ファッショナブル、ゆったり
属性	ユーズド、色、ディテール、定番
装飾	ポケット、ファー付、フリル、リボン
素材	牛革、羽毛、真珠、ウール
ブランド	ディズニー、ユニクロ、シャネル、ティファニー
組み合わせ	色違い、着回し、セット、リバーシブル
機能	品質、弾力、万能、超軽量
商品名	コフレドール、ミリタリーコート、ガウン、リップグロス
形状	Aライン、マキシ丈、スクエア形、オープントゥ
期間	セール、限定、新作、冬物
柄	豹柄、キャラクター、ストライプ、花柄
布	絹、レース、デニム、シルク
色	ネイビー、ホワイト、黒、ベージュ
サイズ	ひざ丈、ロング、ビックサイズ、スモール
対象	お子様、母向け、女性向け、若者向け
地域	米国、日本製、北欧、中国

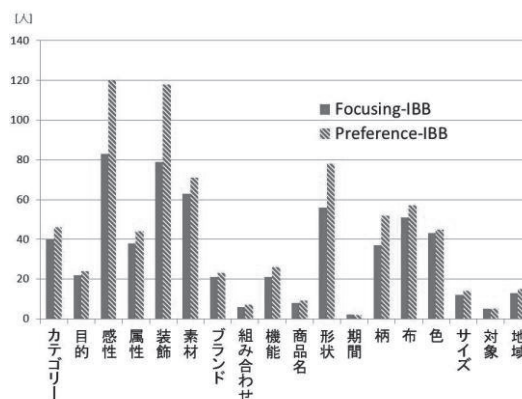


図5 閲覧行動別 ISRWの属性分析

性の分布を示す。このように閲覧行動によって推定される属性の分布は異なり、個人ごとに見てみるとさらに異なるため、閲覧行動の流れを考慮することが有効であると考えられる。

6. まとめと今後の予定

本研究では EC の商品閲覧行動モデルを個人ごとに適用することで、商品閲覧した条件である商品選択条件語 (ISRW) を推定する手法を示した。商品閲覧行動モデルとして Focusing-IBB モデル、Preference-IBB モデルを提案し、それぞれ既存手法と比較し有効であること、それぞれを結合した提案手法も既存手法を上回ることを示した。しかし、提案手法を実装したシステムでは 31% の商品にしか ISRW を推定できないという課題が残っている。その原因としては目的がない Window-shopping のようなランダム性が高い閲覧行動についても対応できるようモデルの改良が必要である。また、今回は用いなかった購買行動についても考慮し、閲覧購買行動を統一的に扱ったモデルの提案をおこない、顧客のニーズ把握やマーケティング戦略に有効となる消費者行動論より深い分析を行っていく。

[文献]

- [1] A. Ahmed, Y. Low, M. Aly, V. Josifovski, and A. J. Smola. Scalable Distributed Inference of Dynamic User Interests for Behavioral Targeting, In Proceedings International Conference on Knowledge Discovery and Data Mining (KDD), pages 373-382, 2011.
- [2] M. W. Berry and J. Kogan (Editors). Text Mining: Applications and Theory, 2010.
- [3] R. W. White, P. N. Bennett, and S. T. Dumais. Predicting Short-Term Interests using Activity-Based Search Context. In Proceedings ACM International Conference on Information and Knowledge Management (CIKM), pages 1009-1018, 2010.
- [4] R. Baeza-Yates and B. Ribeiro-Neto. Modern Information Retrieval. Addison-Wesley-Longman, Harlow, UK, May 1999.
- [5] 阿部誠. 消費者行動理論にもとづいた個人レベルの rf 分析, 日本統計学会誌, pages 239-259, 2008.
- [6] 枝 隼也, 佐藤 哲司. ユーザの Web 探索履歴におけるキーワード遷移に基づく Web ページ推薦システム, 電子情報通信学会データ工学研究専門委員会 他共催, 第 2 回データ工学と情報マネジメントに関するフォーラム (DEIM2010), A3-3, 2010.
- [7] T. Hofmann. Probabilistic Latent Semantic Analysis. In Proceedings Conference on Uncertainty in Artificial Intelligence (UAI), pages 289-296, 1999.
- [8] D.M. Blei, A.Y. Ng, and M.I. Jordan. Latent Dirichlet Allocation. Journal of Machine Learning Research (JMLR), vol.3, pages 993-1022, 2003.

内山 匡 Tadasu UCHIYAMA

昭 60 名大・理・物理卒、昭 62 同大学大学院理学研究科修士課程修了。同年 NTT に入社。平 10-13NTT コミュニケーションズ、平 16-18NTT レゾナントにてポータルサービスの開発等に従事。平 19 より NTT サイバーソリューション研究所主幹研究員。インターネットにおける行動モデリングの研究開発に従事。情報処理学会、電子情報通信学会、日本応用数理学会各会員。

江崎 健司 Kenji ESAKI

日本電信電話株式会社 NTT サイバーソリューション研究所社員。2010 横浜国立大学大学院環境情報学情報メディア環境学修士課程修了。統計科学、機械学習、ネットでの消費者行動モデリングなどの研究・開発に従事。電子情報通信学会、データベース学会各会員。

倉島 健 Takeshi KURASHIMA

2006 年京都大学大学院情報学研究科修士課程修了。同年日本電信電話(株) 入社。現在、NTT サイバーソリューション研究所所属。Web マイニング、行動モデリング、情報推薦の研究開発に従事。日本データベース学会、情報処理学会、電子情報通信学会各会員。

市川 祐介 Yusuke ICHIKAWA

日本電信電話株式会社 NTT サイバーソリューション研究所主任研究員。1996 慶應義塾大学大学院理工学研究科計測工学修士課程修了。グループウェア、通信履歴活用サービス、情報アーキテクチャ、ソフトウェア生産技術などの研究・開発に従事。情報処理学会各会員。

高屋 典子 Noriko TAKAYA

昭 62 年 学習院大学卒平 7 年 NTT 入社、平 7-15 年ネットワークサービスのマーケティング戦略担当、平 16-21NTT レゾナントにてマーケティング部門主幹、平 22 より NTT サイバーソリューション研究所主任研究員。EC における購買行動モデリングの研究開発に従事。日本マーケティングサイエンス学会各会員。