

ネットショッピングにおける商品選択条件の推定

江崎 健司[†] 市川 祐介[†] 高屋 典子[†] 内山 匡[†]

† 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 〒239-0847 神奈川県横須賀市光の丘 1-1
E-mail: †{esaki.kenji,yusuke.ichikawa,uchiyama.tadasu}@lab.ntt.co.jp, ††takaya.noriko@lb.ntt.co.jp

あらまし 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．そのためには顧客を理解するために顧客が閲覧したいと考える考慮集合を把握する事が重要である．そこで，本研究では EC の商品閲覧行動において情報探索のモデルとは異なる行動モデルを提案し，閲覧した商品の商品選択条件語（ISRW）を推定し ISRW に当てはまる商品を考慮集合とすることを考える．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 を推定することである．
キーワード 情報抽出 閲覧ログ 短期的 長期的 商品選択条件 ISRW

Estimating method on choosing the item in E-Commerce

Kenji ESAKI[†], Yusuke ICHIKAWA[†], Noriko TAKAYA[†], and Tadasu UCHIYAMA[†]

† Faculty of Information Science and Engineering, Shuzenji University
1529 Ohira, Izu, Shizuoka, 410-2415 Japan

E-mail: †{esaki.kenji,yusuke.ichikawa,uchiyama.tadasu}@lab.ntt.co.jp, ††takaya.noriko@lb.ntt.co.jp

1. はじめに

EC サイトにおいての重要な課題は，自顧客のニーズの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，顧客の商品選択行動をテーマとした様々な研究が行われている．また，顧客を理解するために，購買結果だけでなく，購買する際候補とした商品（考慮集合）と実際に購買したものの関係を捉え，商品選択行動を明らかにするというアプローチの必要性も唱えられてきた．こうした考慮集合を把握は，実店舗で行うことは難しかったが，EC サイトでは，個人の探索，商品閲覧，購買までを一連の流れとして捕らえることができる．特にファッション，服飾品などの分野では，購買サイクルが不定期であること，同一商品を購買が見られないことなどから商品選択行動の把握が困難であるため考慮集合の把握は重要である．それが実現されることによって考慮集合が形成されていく過程，考慮集合の傾向を把握し，より精度の高い予測やモデリングを行うことが可能になる．そのためにブランドや価格だけで考慮集合の傾向を把握することも可能だが，考慮集合が形成される要因はより多様であり，ある商品が考慮集合として選択された条件をより詳細に把握することが求められている．そこで本研究では，情報探索のモデルとは異なる，EC での商品閲覧

から購買行動への流れを前提として，考慮集合として選択された商品の選択条件語を推定する．以下本稿では，商品選択条件語を ISRW (Item Selected Reason Word) とする．本研究は 2 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のが自然であるため，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．2 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) の流れを考慮して，商品閲覧行動の目的ごとに ISRW を推定する．具体的には，ある条件に当てはまるものを連続して閲覧する行動 Focusing-IBB，単発的に現れるが長期にわたって継続的に出現する顧客の嗜好から起因する Preference-IBB の二つに着眼する．

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

本章では，関連研究として検索クエリと閲覧文章から意図を

推定する手法，商品の特徴や顧客のプロファイルを考慮し商品の推薦を行う協調フィルタ，および時系列を考慮したモデル生成手法に関して述べる．それぞれの手法は，異なるユーザも同一にあつかう場合個人ごとの ISRW 推定に用いることができない，閲覧行動の流れを考慮しない場合閲覧行動の目的ごとの ISRW 推定に用いることができないため本研究に用いることができない．

2.1 文書中から語を推定する研究

検索クエリなどから情報探索の意図を推定する手法や WEB ページや文書中から特徴的な語を抽出する研究 [6][18][13][12] がある．入力した検索クエリと閲覧したページから検索クエリを拡張する手法 [16] やページで頻出する検索クエリをユーザの検索意図として扱う手法が提案されている [20]．また，直接検索クエリの入力がなくとも閲覧したページから重要語を抽出し，時系列で重要語をならべたとき一番関連の高い組を検索クエリとして抽出する手法がある [21]．しかし，これらの研究は全ての閲覧が同一の目的であることを仮定するため閲覧行動の流れを考慮していない「探しているものは何か」を推定する手法は，多くの商品からなぜその商品を選択したのかを推定する ISRW の推定とは必ずしも一致しない．

2.2 協調フィルタを用いた ISRW に関する研究

ある商品に関して，その商品のみた他のユーザが見ている商品も類似した選択条件を持つと仮定してレコメンドする手法として協調フィルタがある．[5]．また各ユーザの興味をプロフィールとして作成し，それを考慮して商品に絞り込む情報フィルタリングの手法を融合させた協調フィルタ手法も多く提案されている [23][22]．潜在クラスでユーザの興味と商品の特徴を表現するモデルを用いた手法や，個人の好みと思われる商品情報を用いた情報フィルタリング手法や対話的遺伝的アルゴリズムで嗜好の変化へ対応する手法などがある．これらの手法は商品の特徴やプロフィールをベースとしているため閲覧行動の流れを考慮していない．

2.3 時系列を用いたモデル生成に関する研究

嗜好を用いる研究としては，ユーザのプロファイルを生成する手法 [1][4][10][9] や閲覧ログを用いて，定期的なドキュメントやページの重要度を推定する手法 [7][8] がある．さらに，短期的・長期的両方を考慮したモデルとして [15][2] がある．[15] は，アイテムを推薦するためにセッションとアイテム，ユーザとアイテムという異なる二部グラフを作成し，セッションで共起したアイテムを短期的，ユーザが同一なセッションすべてで共起したアイテムを長期的な関係とらえランダムウォークを用いたモデルで商品を推薦している．しかし，セッションの中では同一の選択条件を持っているとい仮定しているため閲覧行動の流れを考慮できない．[2] は，LDA (Latent Dirichlet Allocation) を用いた閲覧ログのモデル化を行う際に，ユーザの興味を全体のトレンドに加えて，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線形和で表現することで短期的・長期的な条件をモデル化している．しかし，これらの研究は閲覧行動を一まとまりにして扱うため閲覧行動の流れを考慮していない．

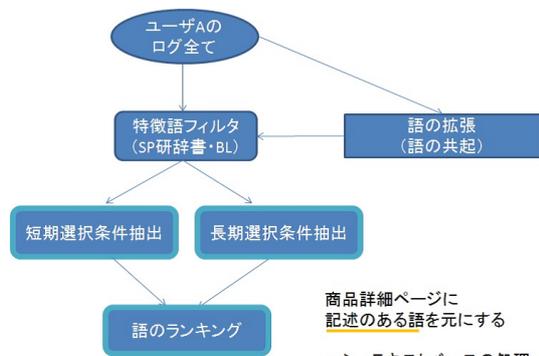


図 1 ISRW 推定システム

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用いた考慮集合として選択された商品ごと個人ごとの ISRW 推定法を提案する．ISRW を個人ごとに推定するためには個人の商品閲覧ログが必要である．なぜなら同じ商品であっても，個人ごとに ISRW は異なり，同じ商品についても同一の個人でも閲覧した時刻，目的により ISRW が異なることがあるからである．本研究ではこのような個人ごと，目的の違いによって ISRW が異なるという認識を前提として ISRW の推定を行う．

本研究で用いる EC サイトの閲覧購買行動ログは約 800 名から収集し，収集期間は約 1 年間 2010 11 月～2011 11 月のデータで収集した対象コマース数は 150 サイトである．本研究ではこのログの中から，商品詳細ページを対象に ISRW の推定を行った．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，商品を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．

図 1 に ISRW 推定システムを示す．まず各モニターが閲覧した商品詳細ページに記述されているテキストに含まれる商品を表現している語 (商品特徴語) を抽出する．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．本文中のテキストに含まれる語では表記ゆれなどの可能性があるため語の拡張を行った．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が高い商品特徴語を用いて行った．この処理によって抽出された商品特徴語の各ページに記述されている商品特徴語語の集合を，以下に述べる各ページの閲覧行動の観点から解析し ISRW を得る．

3.1 商品閲覧行動モデル

ISRW を推定するためには個人ごとの商品閲覧行動の流れ (以下コンテキスト) を考慮する必要がある．商品閲覧する場合に考えられるコンテキストは 2 つあると考えられる．それ以外の場合の多くは "WindowShopping" のような，商品を選択した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るため，ISRW の推定には有用とはいえない．本研究では ISRW の



図 2 IBBSRW の概念図
表 1 記号一覧

Forcusing-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

推定に重要と考えられる 2 つのコンテキスト、すなわち、購入したい商品の条件が明確でそれを集中的に閲覧している場合、Forcusing-IBB、個人の好みにあった商品を閲覧している場合、Preference-IBB、を仮定し、それぞれの商品閲覧行動を考慮した ISRW 推定法を提案する。本研究では提案した各モデルとそれらを結合したモデルそれぞれを用いて算出されたスコアが高い商品特徴語を ISRW として推定する。

Forcusing-IBB は、閲覧したい商品が明確であるため、集中的に閲覧される傾向が高く同一カテゴリで行われることが多い。このような行動においては、異なるページ間においても共通する度合いは高いと仮定し ISRW をモデル化する。一方 Preference-IBB は、自分の好みにあった商品を閲覧しているため、頻度や間隔にはバラつきがみられても、過去から継続して現在まで継続している ISRW が存在すると仮定しモデル化する。

3.2 Forcusing-IBB モデル

Forcusing-IBB は閲覧したい商品の ISRW が明確で同一カテゴリの商品を連続して閲覧すると仮定する。連続して閲覧されたことを同一セッション [14] で閲覧されたと定義し、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。[14]

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

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 Forcusing-IBB モデルのスコアである。 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。図 2 に Forcusing-IBB モデルの概念図を示す。今赤い枠で囲んである商品について Forcusing-IBB を算出する場合を

考える。Forcusing-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、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 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ことを仮定し、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で出現している商品特徴語を対象とする。

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

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は Preference-IBB による選択になる。

この式はマーケティング分野で用いられている生存率という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 [19]。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、Forcusing-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 Forcusing-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。 α はモデル結合の重み係数である。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 商品閲覧行動モデルを用いた個人ごとの ISRW 推定法の実装

Forcusing-IBB モデルで ISRW として推定される候補となる商品特徴語を同一セッション内で 3 商品以上に出現した商品特徴語とした。また、Preference-IBB モデルで ISRW として推定される候補となる商品特徴語を同一セッション内で 3 商品以上に出現した商品特徴語として過去の WEB 閲覧ログに出現したセッション数が 40 回以上である商品特徴語とした。さらに、結合モデルでは明確な ISRW を推定することが重要であるとし、 $\alpha = 0.7$ を用いた。

3.6 精度評価のための比較手法

Forcusing-IBB モデルに対しては同一セッション中での出現頻度をスコアと RAKE 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。

3.6.1 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あつか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。

3.6.2 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次式に示す。[17]

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

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。

3.6.3 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。[17]

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cel1507様が08月23日にE0サイト内の商品詳細ページへアクセスして閲覧された商品を順番に並べたものです。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、十分な候補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候補にない場合は、自由記述欄にキーワードを記入して下さい。

一つ目 (2011-08-23 21:46:34 閲覧)

デニール ボリゴスチル ボウタイ
 フェミニン 4ライン ギャザー
 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 裏地 ブルー
 シフォン パープル 手織き
 アクアガール W4620
 /equipe/

上記以外の表にならなかった理由(キーワードを自由記述)

この商品には覚えがない

図 3 アンケート画面

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

$f(i, w, u)$ 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ログのうち 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、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ったアンケートを行った。各モニターには最大 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。

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 Forcusing-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、Forcusing-IBB が 158 名 388 商品 1666 語、

表 2 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 3 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 %

Preference-IBB が 131 名 191 商品 659 語であった。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計数した。

4.2 評価

本研究では、精度と網羅性という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、 $\text{relevant}(i)$ 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。

まず表 2 に回答された語を分類したものを示す。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、価格が約半数強であることがわかる。また、同一の ISRW によって閲覧された場合、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 3 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 3 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 3 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 3 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 3 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。

これにより本提案システムでは、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。

5.3.2 Focusing-IBB モデル

セッションでの単語の出現頻度や RAKE を用いた手法の精度は 0.42, 0.40 であった。RAKE の特徴は複合語の商品特徴語を重要と扱う点にあるが、ISRW の推定においてはあまり効果がなかったと考えられる。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、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アンケートで回答されるケースが多かったことが原因だと考えられる。さらに提案手法によるカテゴリごとに ISRW を推定するほうが有効であったため精度が改善に有効であったと考えられる。

5.3.3 Preference-IBB モデル

過去に商品特徴語を含む商品の閲覧数を用いた手法の精度は 0.36、過去閲覧した時刻から現在まで時間減衰した商品特徴語を含む商品の閲覧数を用いた手法の精度は 0.45 であり改善した。このことから商品特徴語を含む商品が過去には閲覧され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仮定し、その場合時間減衰を行う方が有効だと考えられる。提案手法も減衰項が入っているため、過去に商品特徴語を含む商品の閲覧数を用いた手法よりも改善したと考えられる。提案手法での精度は 0.49 であり、時間減衰した手法よりも改善した。このことから過去に多く閲覧されているが最近閲覧していない場合よりも頻度は少ないが最近閲覧した商品に商品特徴語を含んでいる場合の方が好みや嗜好が高いと仮定し、その場合提案手法の方が有効だと考えられる。

5.3.4 結合モデル

比較手法は閲覧行動によってスコアを変化させることができないが、提案手法では閲覧行動を考慮することができる。それにより提案手法の方が精度が高く有効であったと考えられる。

5.3.5 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 4 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

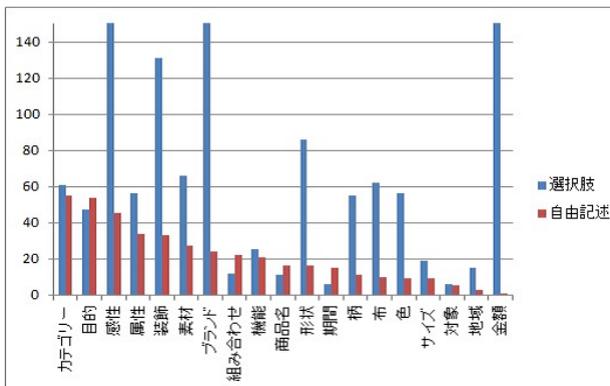


図 4 ISRW の属性分析

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うことがわかった。自由記述によって回答された語は、本研究のシステムでは推定することができないため今後対応していく必要がある。方法としては ISRW として推定する際に商品特徴語で記述できるものを対象とし、機械的に ISRW を推定することが考えられる。例えば、画像やリファラなど関連ページからの商品特徴語の抽出が重要であると考えられる。一方自由記述によって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。このため、商品特徴語で記述できるものを対象にすることで十分であると考えられる。

5.3.6 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-BB と Preference-BB それぞれで推定された属性の分布を示す。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つた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 の商品閲覧行動において情報探索のモデルとは異なる行動モデルを提案し、考慮集合として選択された商品の商品選択条件語 (ISRW) を推定した。Focusing-BB モデル、Preference-BB モデルについてそれぞれ既存手法と比較し有効であることを、それぞれを結合したモデル提案手法で既存手法より有効であることを示した。個人ごとに ISRW が異なることと閲覧行動の目的ごとに ISRW が異なるため、閲覧行動を考慮

表 4 属性分析の例

属性	属性例
カテゴリ	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、若者むけ
地域	米国、日本製、北欧、中国

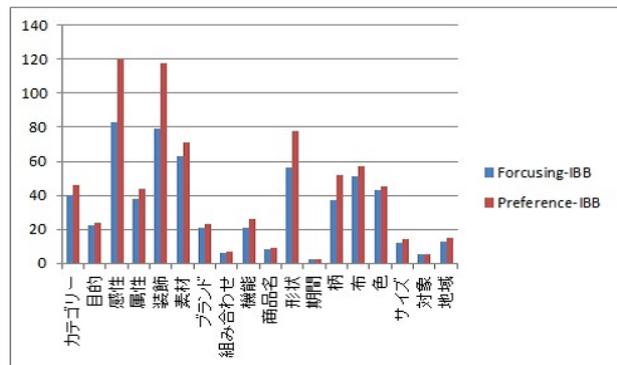


図 5 閲覧行動別 ISRW の属性分析

し個人ごとに ISRW を推定することの有効性を考察した。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。その原因としては商品詳細ページに出現する単語のみを用いるため、商品特徴語が省略されている場合に対応できないからだと考えられる。今後は、ISRW をリファラ・画像から商品特徴語推定しより多くの商品に対して ISRW を推定できる技術に取り組む予定である。また、推定された ISRW に当てはまる考慮集合を用いて精度の高い予測やモデリングを行うという課題がある。今後は、予測やモデリングに有効な ISRW が推定できているのか検証するための研究 [11][3] をおこなう予定である。

文献

- [1] Abdulmohsen Algarni, Yuefeng Li, and Yue Xu. Selected new training documents to update user profile. CIKM '10.
- [2] Mohamed Aly Vanja Josifovski Amr Ahmed, Yudheng Low. Scalable distributed inference of dynamic user interests for behavioral targeting. KDD '11. ACM, 2011.
- [3] Weizhu Chen Gang Wang Qiang Yang Botao Hu, Yuchen Zhang.

- Characterizing search intent diversity into click model. WWW '11. ACM, 2011.
- [4] Paul Alexandru Chirita, Stefania Costache, Wolfgang Nejdl, and Siegfried Handschuh. P-tag: large scale automatic generation of personalized annotation tags for the web. WWW '07. ACM, 2007.
- [5] Brent Smith Greg Linden and jeremy York. Amazon.com recommendations item-to-item collaborative filtering. In *IEEE Internet Computing*, Vol. 7 No. 1, pages pp76-. IEEE, 2003.
- [6] Maria Grineva, Maxim Grinev, and Dmitry Lizorkin. Extracting key terms from noisy and multitheme documents. WWW '09, pages 661–670. ACM, 2009.
- [7] Yuting Liu, Bin Gao, Tie-Yan Liu, Ying Zhang, Zhiming Ma, Shuyuan He, and Hang Li. Browserank: letting web users vote for page importance. SIGIR '08. ACM, 2008.
- [8] Zhiyuan Liu, Wenyi Huang, Yabin Zheng, and Maosong Sun. Automatic keyphrase extraction via topic decomposition. In *Empirical Methods in Natural Language Processing*, 2010.
- [9] Nicolaas Matthijs and Filip Radlinski. Personalizing web search using long term browsing history. WSDM '11. ACM, 2011.
- [10] Feng Qiu and Junghoo Cho. Automatic identification of user interest for personalized search. WWW '06. ACM, 2006.
- [11] Sujith Ravi, Andrei Broder, Evgeniy Gabrilovich, Vanja Josifovski, Sandeep Pandey, and Bo Pang. Automatic generation of bid phrases for online advertising. WSDM '10. ACM, 2010.
- [12] Xiaojun Wan and Jianguo Xiao. Exploiting neighborhood knowledge for single document summarization and keyphrase extraction. *ACM Trans. Inf. Syst.*, 2010.
- [13] Christian Wartena, Wout Slakhorst, and Martin Wibbels. Selecting keywords for content based recommendation. CIKM '10.
- [14] Ryen W. White, Paul N. Bennett, and Susan T. Dumais. Predicting short-term interests using activity-based search context. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, CIKM '10. ACM, 2010.
- [15] Liang Xiang, Quan Yuan, Shiwan Zhao, Li Chen, Xiatian Zhang, Qing Yang, and Jimeng Sun. Temporal recommendation on graphs via long- and short-term preference fusion. KDD '10. ACM, 2010.
- [16] Song Yang and He Li-wei. Optimal rare query suggestion with implicit user feedback. WWW, 2010.
- [17] Ricardo Baeza Yates and Berthier Ribeiro Neto Järvelin. *Modern Information Retrieval*. PEASON, 2011.
- [18] Wen-tau Yih, Joshua Goodman, and Vitor R. Carvalho. Finding advertising keywords on web pages. WWW '06, pages 213–222. ACM, 2006.
- [19] 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. In *日本統計学会誌*, pages 239–259, 2008.
- [20] 高須賀 and 丸山 and 寺田. 閲覧履歴を利用した協調フィルタリングによる web ページ推薦とその評価. In *DEIM*. DEIM, 2010.
- [21] 枝 and 佐藤. 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. pages 431–440. DEIM, 2010.
- [22] 松村, 武田, et al. 選好商品のクラスタリングに基づく嗜好の変化の検出. In *情報処理学会論文誌 Vol.50 No.2*, pages 1234–1239, 2009.
- [23] 麻生, 小野, 本村, 黒川, and 櫻井. 協調フィルタリングと属性ベースフィルタリングの統合について. In *電子情報通信学会信学技報*, pages 55–59. IEICE, 2006.