

複数ショップの利用要因を抽出する行動モデリング手法

高屋 典子^{1,a)} 江崎 健司¹ 石黒 勝彦² 市川 裕介¹

受付日 2015年4月10日, 採録日 2015年10月2日

概要: 本稿では, 商品選択行動の背景にある意図などの要因を抽出するモデル, SHOP-LDA を提案する. 提案法は, 商品の選択過程でのショップ情報を考慮し, ショップ選択と商品選択を同時に説明するトピックを抽出する確率的な行動生成モデルである. 評価実験では, 実ユーザの EC 訪問ログを用いて, 解釈性が高く実用に資する新たな顧客セグメンテーションが可能であることを示した. 本手法は, ID 連携やポイントアライアンスなど, 急速に進展しつつある複数ショップ連動体の共同顧客基盤上で活用しうる, 新たなデータ解析技術であると考えられる.

キーワード: データ解析, パーソナライズ, トピックモデル, EC, 購買行動

The Behavior Generation Modeling Method to Extract the Reason to Select Shops

NORIKO TAKAYA^{1,a)} ESAKI KENJI¹ ISHIGURO KATSUHIKO² ICHIKAWA YUSUKE¹

Received: April 10, 2015, Accepted: October 2, 2015

Abstract: We propose SHOP-LDA, which captures latent patterns of user product adoption behaviors. The proposed model is a probabilistic generative model that extracts “topics”, explaining why users select shops and items. Experimental results with real-world e-commerce user log data show that the proposed model is able to estimate user segments that are interpretable and possibly are beneficial for business practices. Furthermore, this method is applicable to the emerging service to use several shop orchestration such as the ID cooperation and the point alliance.

Keywords: data analysis, personalize, topic model, e-commerce, purchase behavior

1. はじめに

ビックデータの気運の高まりを背景に, 自社に蓄積された顧客データ活用の動きは次第に大きな流れとなっている. オンラインを含む小売流通事業者においても, 市場の成熟化を見据え, 新規顧客の獲得を狙ったプロモーションから, 自社のデータ資産を活かした顧客の囲い込みへと, データ活用の力点を変化させている. 中でも注目されるのは, 単一事業者内にとどまらず, 複数事業者の連合体の中

で, それぞれの保有する顧客資源やデータを相互に活用する営みである [1], [2]. しかし実際には, 顧客はその連合体すべての店舗やサイト (以下ショップという) を利用するとは限らない. 顧客はそれぞれ各自のライフスタイルや商品購入ニーズなどに合わせて利用するショップを選択し行動を行っていることが想像できる. 共同顧客基盤を有効に活用する上でこのような顧客のショップ選択の背景をより深く理解することは非常に重要である. 顧客のショップの選択を理解する上で, 顧客のショップ利用ログを用いてクラスタリングなどを行うことが考えられる. しかしながらこの方法では顧客が現在どのようなショップを併用しているかの実情を知ることができても, どのような商品を探求してショップを訪れているかというニーズも含めて考慮することはできない. そこで本研究では, このような共同顧客基盤を想定し, “誰がどのショップを選んだか” という

¹ NTT サービスエボリューション研究所
NTT Service Evolution Laboratories, NTT Corporation,
Yokosuka, Kanagawa 239-0847, Japan

² NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories, NTT Corporation,
Kyoto, 619-0237, Japan

a) takaya.noriko@lab.ntt.co.jp

ショップ選択行動と“誰がどの商品を選んだか”という商品選択行動双方を考慮し，“誰がどのショップでどの商品を選んだか”という行動ログを用いて，行動の背景となる潜在的な要因を抽出し，相互送客や事業者協調の囲い込み施策につながる新たな解析技術を提案する．提案法は，そのモデル化のベースとして，主にメディア処理の分野で注目されてきたトピックモデル [4] を拡張している．具体的には，商品に加え，新たにショップの選択を考慮し，各顧客（以下ユーザという）のショップの選択と商品（以下アイテムという）の選択を同時に説明可能な確率的行動生成モデルの定式化を提案する．既存のトピックモデルは，ユーザの行動ログから，各ユーザのアイテム選択の要因，すなわちアイテムに対する嗜好を最もよく説明する潜在的な要因“トピック”を抽出し，ユーザの選択を確率的にモデル化する．これに対して，提案法は，ユーザ・ショップ・アイテムの3つの要素のデータを用いて，ユーザの「ショップとアイテム」の選択要因すなわち，行動の意図やニーズを表現する“トピック”を抽出しユーザ行動のモデル化を行う．後述する実験では，複数のECサイト上での閲覧を捕捉したログを用いてモデルを学習し，各ユーザのアイテム閲覧行動の予測精度を計測し，モデルの妥当性を測るとともに，得られたトピックについて解釈を試みる．以下2章では，ベースラインであるトピックモデルの概略と購買などのログへの拡張を試みた既存研究について述べる．3章ではトピックモデルの代表的な例である Latent Dirichlet Allocation (LDA) [3] について概説し，続いてこれをショップとアイテムの選択に拡張した提案法である SHOP-LDA について述べる．4章で，本手法の有効性を評価する為行ったログの収集実験について触れ，5章で，収集実験で実際に得られたログを用いて，提案手法の有効性を定量的，定性的な観点から検証した結果を示す．6章でマーケティングの観点からの有効性を述べ，7章でまとめと今後の展望について述べる．

2. 関連研究

文書などの離散データの解析において，注目されている技術としてトピックモデルがある [3], [4], [5], [6]．トピックモデルは，文書中に含まれる単語の生成過程を確率的にモデル化することで，文書の内容分類やラベル付けすることを実現する手法で，1つの文書を複数の内容，すなわち“トピック”の混合として表現できることを大きな特徴としている．トピックモデルでは，各文書が固有のトピック比率を持ち，文章中のある単語は，トピック比率に従って，ランダムに選択されたトピックを用いてそのトピックに固有の単語分布に従って生成されると表現する．1つの文書を1つのトピックとして扱う混合多項分布に比べ，トピックの表現の柔軟性が高く，音声認識，画像認識 [7]，情報推薦 [8], [9] など多くの分野に適用されている．

トピックモデルは，bag-of-words 形式で表現される様々な離散データに適用可能であるため，メディア処理だけでなく web 閲覧行動 [10]，映画視聴ログ [11]，音楽視聴ログ [12]，購買ログ [13], [14] などへの応用が研究されてきた．文書の中に単語があるという関係をユーザが選択した複数のアイテムの関係に見立て，ユーザの分類やラベル付けへの適用，推薦に用いる方法が提案されている．トピックモデルを用いることのメリットとして，混合多項分布に比べユーザの持つトピック，すなわち興味や嗜好といったものを柔軟に表現できる点があげられる．たとえば，混合多項分布が『私は映画に興味がある』など各ユーザが持つトピックを1つと仮定するのに対し，トピックモデルは『私は映画とワインと旅行に興味がある』という柔軟な表現を可能とする．

トピックモデルでは，先述のトピック比率を用いて各トピックの重要度あるいは強さを定量化することが可能である．つまり映画とワインと旅行，それぞれにどの程度の強い興味を持つか表現できるのである．トピックモデルは教師なしの学習手法であり，興味や嗜好は教師データとして確定的に与えられるものではなく，解析データからボトムアップに抽出される．そのようなトピックモデルの性質を考慮して，Iwata ら [12] は音楽の視聴ログデータを用いて，時間変化するユーザの興味およびアイテムの流行を追跡するためのトピックモデルを提案している．また，購買データ上の各アイテムの価格などの付帯的な情報を用いることによって，より精緻で解釈性に優れたユーザの嗜好を抽出する手法を提案している [13]．これにより同じワインでも，高価格な高級ワインや手軽なテーブルワインを示すトピックを分離することが可能となり，より深いユーザの嗜好理解を可能とした．石垣ら [14] は，価値観やライフスタイルの多様化にともない，年齢・性別・職業などのデモグラフィックの基づく従来のセグメンテーションの困難さから，トピックモデルを行動に基づく大規模なセグメンテーションの方法として位置づけ，購買データからのカテゴリの自動的生成に取り組んでいる．本研究は，ユーザのアイテム選択のログから，背景にある潜在的な要因を抽出するという点では，前述した既存研究と同じ方向性にある．

しかし，これまでの多くのトピックモデルが，文章と単語，ユーザとアイテムなど2つの基本的要素からなるデータのモデル化を試みているのに対し，提案法では，ユーザとアイテムとそのアイテムを選択したショップという3つの要素のモデル化にアプローチしている点で，既存研究と大きく異なる．図 1 に提案モデルの概略を示した．我々はユーザがあるアイテムを選択する際に，複数あるショップの中から1つを選びそのショップを訪れて実際にアイテムを選ぶと仮定した．アイテムの選択過程において，ショップの選好もモデル化するのはごく自然なことであり，またユーザ行動の深い理解のためには，見逃すこと

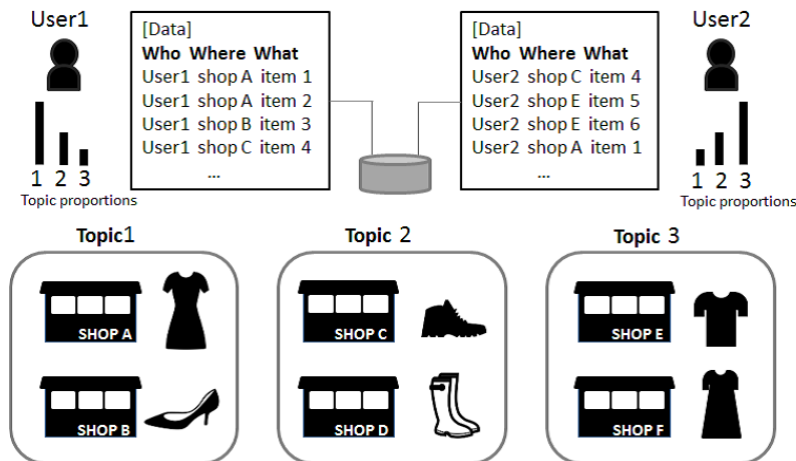


図 1 提案法の概念図

Fig. 1 The conceptual scheme of the proposed method.

のできない重要な要素である。しかし、このようなショップの選好も考慮した商品選択行動^{*1}モデルは、我々の知る限り提案されてこなかった。本提案は3つの要素の確率的モデル化という点でも新たな試みであるが、消費者行動研究 [15], [16], [17], [18] の枠組みにおいても新たな発見と定式化をもたらすものとする。

3. 従来研究と提案手法

本章では、あるユーザのアイテム選択を、ベースラインであるトピックモデルの中でも代表的な手法 LDA で定式化し、次に提案法であるショップ情報を含んだ商品選択行動モデル SHOP-LDA の定式化について示す。

3.1 LDA

Latent Dirichlet Allocation (LDA) [3] は代表的なトピックモデルである。 $u = 1, 2, \dots, U$ はユーザを表し、ユーザが $i = 1, 2, \dots, N_u$ 番目に選択したアイテムを $x_{u,i} = v$ とする。 v はアイテムのインデックスであり、アイテムサイズは V である。また、 N_u はユーザ u が選択したアイテムサイズである。 $z_{u,i}$ はユーザ u が i 番目のアイテムを選択したときに従ったトピックを表し、 $z_{u,i} = k \in \{1, 2, \dots, K\}$ で k 番目のトピックに従うことを表す。総トピック数は K である。また、 $x_{u,i}$, $z_{u,i}$ の集合をそれぞれ $X = \{x_{u,i}\}$, $Z = \{z_{u,i}\}$ とする。

下に LDA の生成モデルを示す。まずそれぞれのユーザ $u = 1, 2, \dots, U$ についてトピック比率 θ_u を生成する。

$$\theta_u \sim \text{Dirichlet}(\alpha), \theta_u \in \mathbb{R}^K \quad (1)$$

続いてそれぞれのトピック $k = 1, 2, \dots, K$ についてアイテムの出力確率 β_k を生成する。

$$\beta_k \sim \text{Dirichlet}(\lambda), \beta_k \in \mathbb{R}^V \quad (2)$$

^{*1} マーケティングや消費者行動論において、あるショップやカテゴリで商品を選択する行動の一般呼称

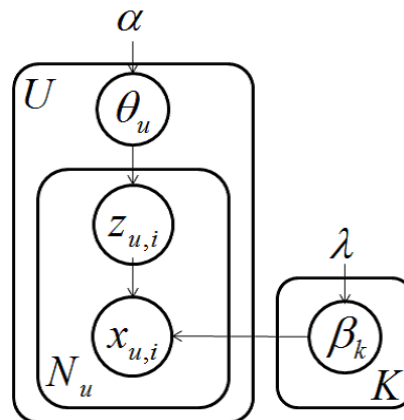


図 2 LDA のグラフィカルモデル

Fig. 2 The graphical model of the LDA.

それぞれのアイテム $n = 1, 2, \dots, N_u$ についてトピック $z_{u,i}$ を生成する。

$$z_{u,i} = k \sim \text{Multi}(\theta_u), k \in \{1, 2, \dots, K\} \quad (3)$$

トピック $z_{u,i}$ と $\beta_{z_{u,i}}$ に従ってアイテム $x_{u,i}$ を生成する。

$$x_{u,i} = v \sim \text{Multi}(\beta_{z_{u,i}}), x_{u,i} \in \{1, 2, \dots, V\} \quad (4)$$

図 2 に LDA のグラフィカルモデルを示す。

トピックモデルでは、各ユーザは、それぞれ固有のトピック比率 θ_u を持ち、アイテムは、この比率に従いトピックを選択したあと、そのトピックに固有のアイテムの出力確率分布 ϕ_k に従って生成されるとしてモデル化する。

このような定式化から、トピックモデルは、ユーザの興味や嗜好などを選択要因からとらえた、柔軟なセグメンテーション技術であるということが出来る。

LDA の推論 (学習) 法は大きく分けて変分ベイズ法 [3] とギブスサンプリング法 [5] に大別することができる。我々の実験ではギブスサンプリング法を用いた。ギブスサンプリングにおけるトピック z のサンプリング式およびパラメー

タ θ , β の事後分布は以下のとおりに計算できる.

$$p(z_{u,i} = k | Z_{\setminus(u,i)}, x_{u,i}, X_{\setminus(u,i)}) \propto \frac{\tilde{\lambda}_{k,x_{u,i}}}{\sum_v \tilde{\lambda}_{k,x_{u,i}}} \frac{\tilde{\alpha}_{u,k}}{\sum_k \tilde{\alpha}_{u,k'}} \quad (5)$$

$$p(\theta_u | Z, X) = \text{Dirichlet}(\tilde{\alpha}_u) \quad (6)$$

$$\tilde{\alpha}_{u,k} = \alpha_k + M_{u,k} \quad (7)$$

$$M_{u,k} = \sum_i z_{u,i,k} \quad (8)$$

$$p(\beta_k | Z, X) = \text{Dirichlet}(\tilde{\lambda}_k) \quad (9)$$

$$\tilde{\lambda}_{k,v} = \lambda_v + N_{k,v} \quad (10)$$

$$N_{k,v} = \sum_d \sum_i z_{u,i,k} x_{u,i,v} \quad (11)$$

3.2 SHOP-LDA

提案法である SHOP-LDA は、ユーザの「ショップとアイテム」の選択を定式化し、ユーザの選択行動の意図やニーズの抽出を行う。SHOP-LDA ではトピックは LDA のようにユーザとアイテムのペアではなく、ユーザ・ショップ・アイテムの 3 つ要素のデータ全体からボトムアップに抽出される。

$u = 1, 2, \dots, U$ はユーザを表し、ユーザが $i = 1, 2, \dots, N_u$ 番目に選択したアイテムを $x_{u,i} = v$ とする。 v はアイテムのインデックスであり、アイテムサイズは V である。また、 N_u はユーザ u が選択したアイテムサイズである。それぞれのアイテムは、それを選んだショップの情報を持つ。これを $y_{u,i} = s \in \{1, 2, \dots, S\}$ とする。 S は全ショップ数である。 $z_{u,i}$ はユーザ u が i 番目のアイテムを選択したときに従ったトピックを表し、 $z_{u,i} = k \in \{1, 2, \dots, K\}$ で k 番目のトピックに従うことを表す。総トピック数は K である。また、 $x_{u,i}$, $y_{u,i}$, $z_{u,i}$ の集合をそれぞれ $X = \{x_{u,i}\}$, $Y = \{y_{u,i}\}$, $Z = \{z_{u,i}\}$ とする。

下に SHOP-LDA の生成モデルを示す。まずそれぞれのユーザ $u = 1, 2, \dots, U$ についてトピック比率 θ_u を生成する。

$$\theta_u \sim \text{Dirichlet}(\alpha), \theta_u \in \mathbb{R}^K \quad (12)$$

続いてそれぞれのトピック $k = 1, 2, \dots, K$ についてアイテムの出力確率 β_k とショップの出力確率 ψ_k を生成する。

$$\beta_k \sim \text{Dirichlet}(\lambda), \beta_k \in \mathbb{R}^V \quad (13)$$

$$\psi_k \sim \text{Dirichlet}(\eta), \psi_k \in \mathbb{R}^S \quad (14)$$

それぞれのアイテム $n = 1, 2, \dots, N_u$ についてトピック $z_{u,i}$ を生成する。

$$z_{u,i} = k \sim \text{Multi}(\theta_u), k \in \{1, 2, \dots, K\} \quad (15)$$

トピック $z_{u,i}$ と $\beta_{z_{u,i}}$ に従ってアイテム $x_{u,i}$ を、トピック

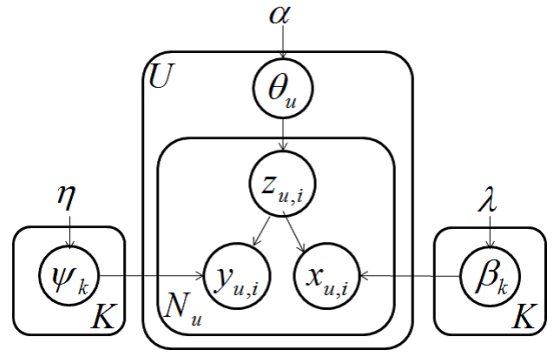


図 3 SHOP-LDA のグラフィカルモデル
Fig. 3 The graphical model of the SHOP-LDA.

$z_{u,i}$ と $\psi_{z_{u,i}}$ に従ってショップ $y_{u,i}$ を生成する。

$$x_{u,i} = v \sim \text{Multi}(\beta_{z_{u,i}}), v \in \{1, 2, \dots, V\} \quad (16)$$

$$y_{u,i} = s \sim \text{Multi}(\psi_{z_{u,i}}), s \in \{1, 2, \dots, S\} \quad (17)$$

図 3 に提案法である SHOP-LDA のグラフィカルモデルを示す。

SHOP-LDA では、各ユーザは、それぞれ固有のトピック比率 θ_u を持ち、この比率に従いトピックを選択した後、各トピックに固有のショップとアイテムの出力確率分布 ψ_k , β_k に従って生成されるとしてモデル化する。提案法が試みていることは、そのユーザのある選択行動において、ショップとアイテムの選択の双方を説明しうる要因の抽出であり、得られたモデルの表現しているトピックは行動の意図やニーズのようなものであるといえる。その意味で、提案法である SHOP-LDA はユーザ各々に特徴的な商品選択行動で表すことのできる、柔軟なセグメンテーション技術であると考えられる。

LDA と同様 SHOP-LDA の学習にはギブスサンプリング法を用いた。ギブスサンプリングにおけるトピック z のサンプリング式およびパラメータ θ , β , ψ の事後分布は以下のとおりに計算できる。

$$p(z_{u,i} = k | Z_{\setminus(u,i)}, x_{u,i}, X_{\setminus(u,i)}, y_{u,i}, Y_{\setminus(u,i)}) \propto \frac{\tilde{\lambda}_{k,x_{u,i}}}{\sum_v \tilde{\lambda}_{k,x_{u,i}}} \frac{\tilde{\alpha}_{u,k}}{\sum_k \tilde{\alpha}_{u,k'}} \frac{\tilde{\gamma}_{k,y_{u,i}}}{\sum_s \tilde{\gamma}_{k,s}} \quad (18)$$

$$p(\theta_u | Z, X) = \text{Dirichlet}(\tilde{\alpha}_u) \quad (19)$$

$$\tilde{\alpha}_{u,k} = \alpha_k + M_{u,k} \quad (20)$$

$$M_{u,k} = \sum_i z_{u,i,k} \quad (21)$$

$$p(\beta_k | Z, X) = \text{Dirichlet}(\tilde{\lambda}_k) \quad (22)$$

$$\tilde{\lambda}_{k,v} = \lambda_v + N_{k,v} \quad (23)$$

$$N_{k,v} = \sum_d \sum_i z_{u,i,k} x_{u,i,v} \quad (24)$$

$$p(\psi_k | Z, X) = \text{Dirichlet}(\tilde{\gamma}_k) \quad (25)$$

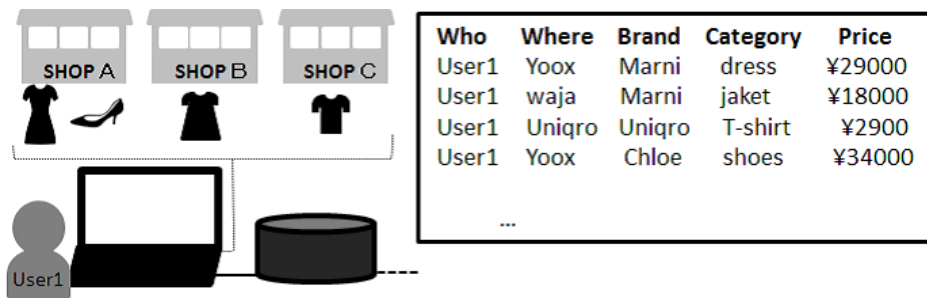


図 4 データ収集システムの概略
Fig. 4 The data gathering system.

$$\tilde{\gamma}_{k,s} = \gamma_s + L_{k,s} \tag{26}$$

$$L_{k,s} = \sum_d \sum_s z_{u,i,k} x_{u,i,s} \tag{27}$$

4. 実験準備

ユーザのアイテム閲覧や、購買のデータは、一般的には単一のショップのものが多く、しかし本研究の目的は、複数ショップの商品選択行動を横断的に取得し利用する共同顧客基盤上で有効な、新しいデータ解析技術を考案することにある。共同顧客基盤上で動作するグループウェアとしての解析技術の提案である。このため、提案手法の有効性を評価するために、想定する環境に近い状況で評価を行う必要があると考えた。

4.1 行動ログ収集システム

今回用いたログ収集システムは、PCブラウザを用いたインターネット利用、とりわけECサイトの利用ログを収集するものである。図4にそのデータ収集システムの概略を示す。具体的には、ユーザのアクセスしたURL情報とECサイト利用時の閲覧内容を取得可能な専用ソフトウェアを用いた。参加するユーザは、このソフトウェアをダウンロードして、ECサイト利用時にこれを起動する。ソフトウェアは、ログインによってユーザを識別するIDを発行する。取得されたURL情報は、URLと対応つけられたECサイトのマスターデータでECサイト名に置換される。アイテム情報に対しては、取得されたタイトル、詳細説明文などに言語処理を施し、カテゴリ、ブランド名、価格などの属性情報を抽出する。この処理には、NTTメディアインテリジェンス研究所の言語処理技術リッチインデクサ^{*2}を用いた。抽出された情報は、各アイテムのユニークなIDとともに、データベース上に登録される。

4.2 ログ収集実験

EC利用ログの収集は、2010年11月から約2年間にわたってファッション、コスメの閲覧行動を対象に行った。

^{*2} http://www.ntt.co.jp/svlab/activity/category_2/product2.07.html

ファッションコスメは、女性ユーザのECサイト利用の上位を占めるカテゴリであり、マーケティングにおいても横断的分析がなされてこなかったカテゴリである。参加ユーザは実験の主旨に賛同し、参加を申し込み、プライバシー情報の取得に同意した上で、ソフトウェアのインストールを行う。実験の性質としては、ネット視聴率調査会社のオペレーションと同様の方法である。ソフトウェアには、取得データをサーバに送信する前に確認する機能が備わっており、ユーザ自身が自ら確認を行ったうえで、送信する内容の開示制御を細かく行うことができる。ユーザにはあらかじめ、ログ収集の対象となるECサイト名を公表しており、そのECサイトを利用する際にソフトウェアの起動を依頼し、ログを取得した。実験参加者は約1,431名、実際に利用されたECサイトは150、閲覧されたブランド数は10,043に上る。対象とした150サイトのうちよく訪れられていた上位50ECサイトのリストを表1に示している。当初多くのユーザが、国内ECサイトのみを利用していることを想定していたが、表2のように2011年時点で欧米や中国などの海外ECサイトも積極的に利用していた。

5. 評価実験

前章で述べたログ収集実験で得たデータを用いて、提案法であるSHOP-LDAの有用性の評価を行った。評価は定量的、定性的な以下の2つとした。

- (1) 定量評価：学習期間のデータを用いてモデル学習を行い、テスト期間の行動を予測してその精度を評価する。
- (2) 定性評価：従来法であるLDAと提案法で得られた潜在的要因（トピック）を比較する。

本章では、まず評価実験に用いたデータセットについて述べ、次いで2つの実験結果を示し考察を行う。

5.1 評価実験用データ

評価実験には、ログ収集実験の最後期である2012年10月から12月のユーザのECサイト利用とアイテム閲覧データを用いた。評価実験では、ユーザIDから得られるユーザ識別情報、URLの変換から得られたECサイト情報、そして、アイテムの代わりとして、ブランド情報を用いた。

表 1 ユーザ利用数上位 50 ショップ

Table 1 Top 50 sites in terms of the user utilised frequency.

ショップ名	利用者数	総ブランド数	閲覧総数
楽天市場	1,323	7,173	335,277
Amazon	1,180	2,324	55,102
Yahoo!ショッピング	1,034	2,256	41,206
Yahoo!オークション	830	4,259	171,178
ELLE SHOP	804	537	19,805
ZOZOTOWN	694	1,752	64,526
ユニクロ	577	2	8,199
バイマ	405	519	3,097
ニッセン	382	46	3,097
無印良品	371	1	6,801
DHC	316	6	859
マガシーク	300	489	10,951
ワールド DIRECTSTYLE	285	108	6,086
集英社 FLAG SHOP	281	382	3,149
OIOI	277	463	3,272
ジャバリ	264	433	5,207
セレクトソニック	251	74	6,984
フェリシモ	246	47	1,452
ミラベラ	243	184	4,742
オルビス	234	6	857
i ルミネ	230	229	8,051
ワコール	223	34	2,542
伊勢丹	222	337	502
ガールズウォーカー	217	164	2,457
ユニテッドアローズ	210	124	3,405
ドクターシーラボ	207	1	316
YOOX	200	902	4,525
高島屋	199	183	1,915
スタイルクルーズ	194	352	7,879
セレクトスクエア	186	110	5,778
FOREVER21	154	3	4,524
オンワード CROSSET	151	37	2,270
ファッションウォーカー	146	163	1,470
BARNEYS NEW YORK	130	136	1,380
アウトレットピーク	130	407	7,273
waja	121	222	3,234
CHANEL	109	1	23
サンヨー・アイストア	105	48	3,254
イケセイキレイ	98	72	73
ブランデリ	90	359	6,288
Shopbop	79	340	3,572
LANCOME	75	2	65
NET-A-PORTER	75	257	2,125
ルック・イーショップ	75	14	537
Neiman Marcus	70	149	337
セレクトカフェ	68	88	1,275
ゲンゼ	64	4	574
A-net	62	24	1,456
American Apparel	55	4	493
VICTORIAS SECRET	53	19	353

表 2 ユーザ利用数上位 5 ショップ (海外)

Table 2 Top five sites in terms of the user utilised frequency at the overseas.

ショップ名	訪問者数	総ブランド数	閲覧総数
YOOX	200	902	4,525
SHOP BOP	79	340	3,772
NET-A-PORTER	75	257	2,125
Neiman Marcus	70	149	337
THE OUTNET	24	217	1,163

表 3 上位 5 件出力時の予測正答率 (割合)

Table 3 The predicted performance of 5 best prediction.

	LDA	SHOP-LDA
トピック数 3	25%	23%
トピック数 5	26%	25%
トピック数 7	25%	30%
トピック数 10	30%	30%

EC サイト (以下ショップという) 上の商品の多くは, JAN コードなど共通の商品コードを持たないため, アイテムを一意に定義することが難しい. このため, 商品のタイトルや説明文から言語処理によって抽出されたブランド名をその代替とした. 期間中閲覧を行ったユーザ数は 514 人, 利用されたショップ数は 69, 閲覧されたユニークブランド数は 3,925, 総閲覧数は 65,073 になる.

5.2 定量的な評価

定量的評価では, 学習期間として 2012 年 10 月から 12 月の最終週を除くデータを利用して学習を行い, 12 月の最終週の各ユーザの閲覧を推定するという方法をとった. 評価尺度として N ベスト正答率を用いた. N ベスト正答率は, モデルが出力した上位 N 件の中に, 実際に閲覧されたものが含まれている人数の割合である. 個人の過去の閲覧ログを学習データとして次期間内のユーザの閲覧を予測し, 何人の閲覧を予測できたか評価する. 12 月最終週に閲覧を行ったユーザ数は 210 人, 閲覧された総ブランド数は 1,241 であることをふまえ, 各ユーザに対して推定する数を 5 とした. 表 3 に提案法である SHOP-LDA と従来法である LDA の正答率を示す. パラメータであるトピック数を 3 から 10 まで変化させそれぞれの正答率を記載した.

両手法とも, 評価期間のユーザ数 210 人に対して, 正答率は 20% から 30% つまり 40 人から 60 人程度に対して適切な推定ができたことを示している. 仮に対象期間内に 5 ブランドをユーザに提示したとして, 20% から 30% のユーザにとって妥当な提示となる. 提案手法である SHOP-LDA は, 予測の精度の面では, 従来法である LDA を大きく上回ることはなかった. この理由として, 提案法がショップとアイテムの双方を最適に説明する学習を行うのに対して, LDA はアイテムのみに最適化した学習を行っていることがあげられる. LDA がトピック数を上げるにつれ, 推定精度を上げていくのに対して, 提案法ではトピック数 7 で推定精度が 30% を超え, 10 トピックにしても推定精度が横ばいになる. 提案法において, トピック数 7 では, LDA の推定精度を 5% 上回った. このことから, 今回のデータでは予測の正答率という面から考えて, 妥当なトピックの数は, 7 程度であると推察できる.

図 5 にトピック数を 7 に固定し, LDA と SHOP-LDA の上位 N 件を出力した際の正答率を示した. X 軸が変化

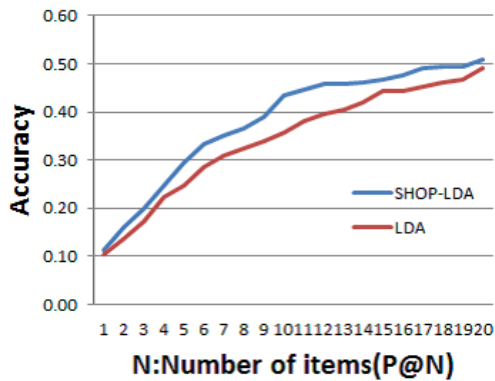


図 5 上位 N 件の予測正答者数

Fig. 5 The predicted performance of N best prediction.

表 4 上位 5 件のショップの出力時の予測正答率 (割合)

Table 4 The predicted performance of 5 best prediction.

	LDA	SHOP-LDA
トピック数 3	67%	66%
トピック数 5	69%	70%
トピック数 7	70%	70%
トピック数 10	71%	70%

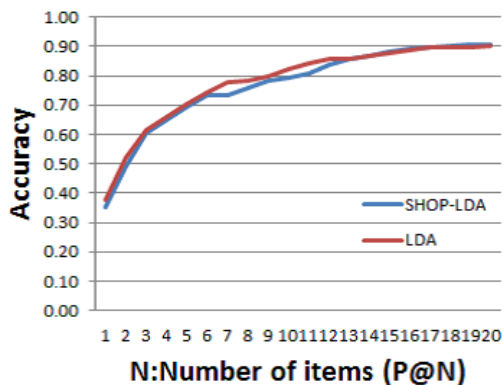


図 6 上位 N 件の予測正答者数

Fig. 6 The predicted performance of N best prediction.

させた N の数, Y 軸が予測正答率である. 上位 1 件のブランドを出力した際の正答率は 11%, 上位 5 件で 30%, 上位 10 件で 40%, 上位 20 件で 50% を超え. LDA と比較して, 高い正答率を得られた.

次に提案法がショップの選択も説明可能なモデルとしても妥当であることを確認するために, 提案法におけるショップの予測精度を確認した. パラメータであるトピック数を 3 から 10 まで変化させた正答率を表 4 に記載した. 参考値として, LDA を用いてショップの選択に適用した際の値を左に記載している. アイテムと同様にトピック数を増やすことで正答率は上昇し, トピック数 7 で 0.7 すなわち全ユーザーの 70% のテスト期間の利用ショップを予測できていることが分かる. 図 6 に, トピックを 7 として N の数を 1 から大きくした際の正答率の変化を記載した. 参

考値である LDA の適用結果と比較しても, 同等の予測正答率を示しており, 提案法がショップの選択も説明可能なモデルであることが裏付けられた. 以上のように, 提案法はショップとアイテム選択を同時に説明するトピックを抽出し, ユーザの次の行動を推定する手法として有効であると考えられる.

5.3 定性的な評価

定性的な評価として, 従来法と提案法で同じトピック数 7 で, LDA では得られたブランドの上位 10 件とショップの上位 10 件を, SHOP-LDA では, ショップ, アイテムの上位 10 件ずつを出力して, トピックの解釈を行う. 2 つのモデル, LDA ではユーザーのアイテム選択の潜在要因, SHOP-LDA ではユーザーのショップとアイテムの選択要因をどのように抽出したかについて考察を行う.

5.3.1 LDA のブランドトピック解釈

LDA の出力した各トピックでの上位 10 件のブランドを表 5 に示した. 左から右にトピック番号, 上から下にブランド名を記述している. トピック番号は学習時に割り振られていく番号であるため数字に意味はない. 各トピックのブランド列は上に記載されているほど, そのトピックで出現する確率が高いことを示しており, 各トピックの上位を意識して解釈する必要がある. 表の最下部に各トピックの定性的な解釈を付記した. トピック 1 は, コスメ, インテリア雑貨が多く, 中でも 3 番目と 4 番目のキャス・キッドソンとクォーターリポートに特徴が表れている. 両者はそのテキスタイルの魅力で人気があり, クォーターリポートはカーテン, テキスタイル小物が人気である. キャスキッドソンは特徴的なテキスタイルで展開された小物, バックなどで, 当時日本上陸直後でもあり大人気を博した. これと対照的なのが, トピック 6 のユニクロ, 無印良品である. 両者はシンプルで限りなく無地に近い. さらに対象的なトピックとして, トピック 7 があげられる. 上位ブランドを見ると, シンボリックなキャラクターやシグニチャーと呼ばれるシンボルマークを持つ. 画像処理を施すのではなく, このようなブランドの持つ特徴的な魅力が抽出できている点が注目される. また隠れた嗜好をトピック化するという意味では, トピック 5 は, ユニセックスでスポーティブなクラスである. トピック 5 はコスメ系が強く, トピック 2 はオンラインセレクトショップで人気の国内ブランド, トピック 3 はインポートハイブランドが同じクラスとして抽出されている. このように, LDA はユーザーとアイテムの 2 つの要素から嗜好を抽出する.

5.3.2 LDA のショップトピック解釈

次に, 前項で行ったショップの LDA によるショップの選択のモデル化によって得られたトピック 7 つとその上位 10 ショップの結果を, 表 6 に示す. トピック 1 は, Yahoo!オークションを中心とするトピックで, このトピックには,

表 5 LDA で得られたトピックと上位ブランド

Table 5 The topic and top branding item obtained by the LDA.

topic 1	topic 2	topic 3	topic 4	topic 5	topic 6	topic 7
g.u. DHC キヤス キッドソン クォーターリポート バナソニック Ray Cassin ヴェルニカ イブサンローラン ヘリーハンセン シャネル 服飾小物(バック・靴) コスメ・美容機器 インテリア雑貨	ドゥーズイエム クラス マルティニーク ガリヤルダ ガランテ BEAUTY&YOUTH WOMENS nano・universe Ladies ROSE BUD CIAOPANIC イエナ MACKINTOSH PHILOSOPHY green label relaxing 国内セレクトショップ系 ファッション	シャネル duvetica CHLOE' LOEWE marni エルメス tommy hilfiger MAISON MARTIN MARGIELA 22 ジュシークチュール ゴールデンゲース インポートハイブランド 服飾小物(バック・靴) “最新トレンド”	ugg プーマ Another Edition アディダス byステラマッカートニー NIKE ラルフローレン PHILOSOPHY DI ALBERTA FERRE UNIQLO アルベルタフィレットィ Reebok 服飾小物(バック・靴) “スポーツカジュアル” “ナチュラル”	nissen フォーエバー21 ドクターシーラボ comme des garçons denim dungaree zoetie kriff mayer humor スナオ クワハラ ワコール コスメ キッズ インナー インテリア雑貨	UNIQLO 無印良品 ヒューマン ウーマン g.u. Foxey ludlow ヴァンクリーフ&アーベル レスポートサック ジル スチュアート ケイト・スペード NY ファッション 宝飾・服飾小物(バック) ファストハイブランド混合	disney Coach ネネット X-girl エミュー グッチ ミキハウスファースト くまのプーさん ラルフローレン アンパンマン ファッション インテリア雑貨 “キャラクター”

表 6 LDA で得られたトピックと上位ショップ

Table 6 The topic and top branding item obtained by the LDA.

topic1	topic2	topic3	topic4	topic5	topic6	topic7
Yahoo!オークション Yahoo!ショッピング Neiman Marcus Amazon ワコール A-net 高島屋 GUCCI 7netshopping ファミリア	ユニクロ ブランデリ Amazon BARNEYS NEW YORK ニッセン DHC 無印良品 レモール Yahoo!ショッピング ハーバー	ZOZOTOWN スタイルクルーズ セレクトスクエア マガシーク Amazon ミラベラ ユナイテッドアローズ アウトレットピーク ALOHA RAG CHANEL	Yahoo!ショッピング アウトレットピーク マガシーク ワールドダイレクスタイル 集英社 FLAG SHOP ファッションウォーカー セレクトソニック ワコール ユナイテッドアローズ Shopbop	楽天市場 waja Amazon LANCÔME ショップチャンネル ルック・イーショップ Neiman Marcus GUCCI Shopbop ハーバー	ELLE SHOP 無印良品 Amazon ドクターシーラボ ジャパリ ミラベラ 伊勢丹 ダイアナシューズ LANCÔME 高島屋	YOOX サンヨー・アイストア FOREVER21 楽天市場 NET-A-PORTER THE OUTNET A-net VICTORIAS SECRET American Apparel ショップチャンネル

赤ちゃん本舗を持つ7netshoppig や子供服ブランドファミリアなどが含まれるが、同時に Neiman Marcus や高島屋 GUCCI などの高価格帯を販売するサイトも抽出されている。トピック 2 においても、ユニクロや無印良品と通販サイトニッセン、レモール、総合 EC であるアマゾンと Yahoo!ショッピング、DHC やハーバーなどのコスメを中心としたショップが混在して表れている。トピック 3 と 4 は、女性のファッションを中心としたトピックだが、トピック 5, 6, 7 で総合 EC とファッション、コスメ、シューズなどのショップが混在する。ショップ利用予測の正答率では、やや高い精度を持つ LDA であるが、得られたトピックにおいては、1つのトピックに様々なショップが混在し、解釈可能な傾向を得ることが難しい結果となった。2章で述べたとおり、トピックモデルを用いるメリットとして、あるユーザのトピックの比率を記述することで、そのユーザがどのような傾向を持ったユーザか理解できる点があげられる。しかし、比率が記述されてもトピックが解釈可能でないとそこから知見を見出すことはできない。ショップ情報だけ用いてトピックモデルを適用しても、ショップの選択行動の背景を導くには不十分であると考えられる。

5.3.3 SHOP-LDA のトピック解釈

LDA がユーザの嗜好を表現するのに対して、SHOP-LDA はユーザの商品選択行動の要因を抽出すると仮定した。アイテム選択行動の要因とは、ユーザがどのような理由でそのショップに行きどのようなアイテムを見るのかという、行動に至る意図、あるいはニーズのようなものである。定性的な考察を行うトピック数は、定量評価で正答率の高かった 7 とする。結果を表 7 に示した。最下部に各トピ

クの解釈を示した。

まず特徴的なトピックとしてあげられるのが、トピック 2 の Yahoo!オークション、楽天市場、Yahoo!ショッピング、アマゾン、伊勢丹で、シャネル、フォクシー、ヴァンクリフ・アンド・アーベル、エルメスなどのハイブランド、特にバックや宝飾品などの服飾小物を見るクラスである。ハイブランドは単価が高く新品を買うのは高価であるため、オークションで中古品を、あるいはオンラインモールで並行輸入品を選ぼうとするという行動がみられる。トピック 2 はこのような行動を端的に表現しているものと考えられる。トピック 7 は、同様にハイブランドのクラスであるが、海外 EC サイトを訪問しアイテムを探すというクラスが塊となって表れている。これらのショップは、海外からの発送になるため送料もかかるが、日本に入ってきていないアイテムを手に入れられることが魅力である。ヨーロッパ、アメリカのほか、個人バイヤが海外から買い付けたアイテムを販売する waja がこのトピックに属している。これらのショップは、オンラインセレクトショップと呼ばれ、各ショップのバイヤが、各バイヤの目線ならではの仕入れを行っている。このトピック比率を多く持つユーザは、こうした各店のバイヤの行うセクション、つまりアイテムラインナップを楽しむためにショップを訪れていると考えられる。本手法は、ユーザの隠れたショップとアイテムの選択行動の要因、意図やニーズを同じトピックとして抽出することに有効であることを示しているといえる。

トピック 7 が海外オンラインセレクトショップであるのに対し、トピック 6 には国内のオンラインセレクトショップが表れている。このクラスでは、マーク・バイ・マーク

表 7 SHOP-LDA で得られたトピックと上位ショップおよびブランド
Table 7 The topic and top branding item obtained by the SHOP-LDA.

topic 1	topic 2	topic 3	topic 4	topic 5	topic 6	topic 7
楽天市場 無印良品 A-net Yahoo!ショッピング ルック・イーショップ ハーバー LANCOME CHANEL MatchesFashion RALPH LAUREN	Yahoo!オークション 楽天市場 Yahoo!ショッピング Amazon 伊勢丹 GUCCI オルビス 高島屋 Neiman Marcus CHANEL	ZOZOTOWN ユニクロ スタイルクルーズ ユナイテッドアローズ FOREVER21 Woolool iルミネ ジャパヴ ファッションウオーカー A-net	マガシーク セレクトソニック ワールドダイレクトスタイル サンヨー・アイストア 集英社 FLAG SHOP セレクトスクエア ファッションウオーカー ONWARD CROSSSET OIOI iルミネ	Amazon 楽天市場 Yahoo!ショッピング ドクターシーラボ ニッセン DHC フェリシモ VICTORIAS SECRET ジャパヴ OIOI	ELLE SHOP ミラベラ 集英社 FLAG SHOP BARNEYS NEW YORK ジャパヴ ZOZOTOWN 伊勢丹 ANAP iルミネ ルック・イーショップ	YOOX ブランドリ Shoptop NET-A-PORTER THE OUTNET Neiman Marcus waja スタイルクルーズ Yahoo!ショッピング ミラベラ
topic 1	topic 2	topic 3	topic 4	topic 5	topic 6	topic 7
無印良品 イッタラ nissen duvetica ワコール シャネル zootie マリメッコ アネモネ デオール インテリア雑貨 服飾小物 コスメ・インナー	シャネル ワコール Foxey comme des garçons ヴァンクリーフ&アーペル エルメス denim dungaree LOEWE ジル スチュアート CHLOE インポートハイブランド 服飾雑貨(バック・靴) 宝飾・アクセサリー	UNIQLO g.u. フォーエバー21 ドゥーズィエム クラス BEAUTY&YOUTH WOMENS ROSE BUD nano・universe Ladys CIAOPANIC TYPY green label relaxing X-girl ファッション "ファスト&カジュアル"	ヒューマン ウーマン TO BE CHIC ストロベリーフィールズ NARACAMICIE エポカ FREE'S MART ジャスグリッティ ジル スチュアート MACKINTOSH PHILOSOPHY ICB ファッション ライセンスブランド "通勤・コンサバティブ"	ドクターシーラボ nissen レスポートサック coach Gloverall DHC NIKE ビクトリアズ シークレット American Apparel アディダスbyステラマッカート コスメ・美容機器 服飾小物・インナー "スポーツティブ"	ミュベール ガリヤルダガランテ ludow アルファエー マルティニーク マークBYマークジェイコブス MACKINTOSH PHILOSOPHY ケイト・スペード NY SACRA ジル スチュアート 海外セレクトショップ ファッション ライセンス&TOKYOブランド	トリーパーチ Marni PHILOSOPHY DI ALBERTA FERRE シーバイクロエ CHLOE Ray Cassin Carven アルベルタフィレッティ ダイアン ファーステンバーグ ステラ マッカートニー 海外セレクトショップ ファッション ハイブランド

ジェイコブスやケイト・スペード NY などの現在国内で人気のあるライセンスブランドのほかに、海外から新たに注目されている TOKYO ブランドが属している。国内のブランドショップの中では、ファッション界の最先端トレンドでセクションを行う、ELLE SHOP, FLAG SHOP など雑誌社のオンラインショップが同じクラスとして抽出された。さらに、トピック 4 は国内セレクトショップでも、ふだんの通勤通学などに好まれる服を取りそろえたショップとブランドを抽出している。ワールド、三陽商会など国内ファッションメーカーがこのクラスに属している。類似したトピックとして、ZOZOTOWN, ユニクロなどを含むトピック 3 があげられる。トピック 3 の代表的なブランドはユニクロ, GU, フォーエバー 21 となり、トピック 4 よりさらにカジュアルで安価なブランドが属している。トピック 5 とトピック 1 は、ファッション雑貨インテリア, コスメの多いトピックであるが、代表的なブランドには価格帯の異なりがある。

以上のように、提案法である SHOP-LDA は、ユーザの EC サイト訪問とブランド閲覧履歴から、ユーザのアイテム選択行動の意図やニーズのようなものを抽出することに成功していると考えられる。

5.4 2 手法のトピック比較と考察

この項では、これまで述べてきた 2 つの手法が抽出したトピックを比較して考察を行う。国内ファストファッション市場の代表的ブランドであるユニクロ, GU, フォーエバー 21 は、LDA では異なるトピックに割り当てられた。しかし、SHOP-LDA では、これらを同じトピックとしてとらえることができている。Iwata ら [13] が、価格帯を明示的にトピック形成の要因としてモデルの中に組み込んでいるのに対して、SHOP-LDA はショップの情報を入れることによって、ブランドの価格的な違い含んだトピック抽出に成功している。各ショップは、その取り扱うカテゴリ

(靴・バック・コスメなど)に加えてテイストや価格帯に個性を持つ。SHOP-LDA はショップの選好を同時にモデル化することによって、カテゴリ・テイスト・価格帯などで特徴づけられるブランドのクラス抽出を実現していると推察される。

一方 SHOP-LDA において、利用者や品揃えに幅のある楽天市場や、Yahoo!ショッピングなど、総合モールは、トピック 1, 2, 5, 6 などのように、複数のトピックに分かれて現れている。LDA では、楽天市場や、Yahoo!ショッピングは、2 つのトピックに所属していたが、ブランドの情報をういた SHOP-LDA では、それぞれ 3 つに分けることが妥当と抽出された。逆に、LDA で 4 つのトピックに所属していたアマゾン、高級ファッションブランドは取り扱っていないなどブランドの品揃えの幅が狭いため、ブランド選択も同時に説明する提案法では 2 つのトピックにしか出現しない。SHOP-LDA では、同じ EC サイトを利用してユーザであっても、実際にそこで見ているブランドを考慮することで、何を目的に EC サイトを訪れているかという意図の違いを考慮することが可能になり、単にショップ利用ログを解析しただけでは明らかにできないユーザとショップの関係を表現できる。このように SHOP-LDA では、単にアイテムの傾向ではなく、ユーザの行動の意図やニーズを描出しながら、ショップとブランドをクラス化し、各ユーザのショップとアイテムの選択行動をモデル化することができているといえる。

6. マーケティング施策への適用性

最後に、評価実験結果をもとに、提案法である SHOP-LDA のマーケティング施策導出における有効性について述べる。本研究の狙いは、複数ショップの相互送客や、顧客囲い込みの施策立案において活用可能なデータ解析技術の創出であった。以下、この観点から評価実験結果を考察してみたい。まず、定量的な側面では、各ユーザ上位 5 件

のブランドの出力で、おおよそ 30% のユーザの翌週の閲覧を予測することができた。各ユーザの 1,241 ブランド中からの選択行動が、モデルの出力する上位 5 つで説明できていることから、実務に耐えるモデルであると考えられる。

続いて定性的な評価結果の考察をする。本研究の主眼が、マーケタの戦略立案支援を目的としていることから、定性的な解釈性は非常に重要である。今回の提案法を用いて導かれた出力結果によって、同じトピックに属している事業者は、ユーザの同じようなニーズを満たしていると考えられる。すなわち、行動の意図やニーズの点からみて同じセグメントのユーザを共有していることになる。各ショップが同セグメントのユーザニーズに訴求しているという事実は、仮に同じトピック内の複数ショップが連携した場合に、他のトピックとの差異を明確化しうる。より強固な顧客基盤を獲得できると解釈できる。しかしその反面、同じトピックに属する事業者は、ユーザの同じニーズを満たしているという点で競争関係にあるとも考えうる。たとえば、トピック 6 の ELLE SHOP (ハースト婦人画報社) と FLAG SHOP (集英社) は、『雑誌に取り上げられている最先端のファッションを手に入れたい』とユーザニーズに込めている点で、同じ価値を提供しているショップであり、互いに競い合う関係にあるともいえるのである。両者が相互送客を行うことは、売り上げの点からみると機会損失となる可能性もある。一方これらのショップにとって、シューズに特化して豊富な品ぞろえを持つ、ジャバリとの協業は、シナジーを生みうる相互送客パートナーシップが実現できるかもしれない。

次に、各ショップに商品を提供するブランドホルダの視点で結果を考察する。国内で高い人気を博すジル・スチュアート、マッキントッシュ・フィロソフィーは ELLE SHOP、ミラベラなどトピック 6 のセレクトショップの他に、マガシークなどのトピック 4 のショップにも商品を提供している。つまり、2 つのブランドは、トピック 4 と 6、2 つのユーザセグメントに訴求できているといえる。このような場合、たとえば自社商品の中でも、最先端でエッジの効いた、シンボリックな商品は、ELLE SHOP、ミラベラで販売し、ふだんの通勤などでも着ることができる商品をマガシークに供給するなどのマーチャライジング施策が考えられる。訴求するターゲットの違いを、より考慮した販売チャネル戦略の立案が可能となるのである。SHOP-LDA は、ユーザの商品選択行動の背景にある意図やニーズを抽出し、この観点からユーザをセグメント化できる技術である。またこのような知見は、データ解析の手法としてだけでなく、マーケティングにおける消費者行動論の枠組みにおいても新しい知識抽出手法として評価できるものと考えられる。提案法は、ショップやブランドホルダなど、様々な事業者の戦略立案に有効なデータ解析技術であると確信している。

7. まとめと今後の課題

本稿では、トピックモデルの代表的な手法である LDA をベースに、ユーザがどのショップでどのアイテム閲覧したかという 3 つの要素で構成されるデータから、潜在的な顧客セグメントやユーザの選択行動の背景にある意図やニーズを抽出するモデル、SHOP-LDA を構築した。

また実際の行動ログを用いた評価実験を行い、より解釈性の高く実用に資する新たな顧客セグメンテーションが可能であることを示した。さらに、定性的な評価だけではなく、構築したモデルによって閲覧行動を予測し、高い行動予測精度を得られることを検証した。複数ショップの共同顧客基盤において、本手法がショップの選好を考慮したより深いユーザ理解を可能とする技術の導出であり、複数ショップでの相互送客やアライアンスパートナーシップ戦略の策定に資する方法であること実証をできた。今回の実験では、行動予測の精度としては、従来法を大きく超えることはできなかったが、得られたトピックを活用して、より精度の高いモデルを構築し、複数事業者間の横断的レコメンデーションなど活用性の高いアプリケーションを目指していきたい。

参考文献

- [1] exchange wire Japan: Yahoo!JAPAN が仏クリテオ社と提携、初の第三者配信広告開始 (オンライン), 入手先 <<http://www.exchangewire.jp/2012/08/31/yahoojapan-criteo/>> (参照 2012-08-31).
- [2] IT pro: ソフトバンクが T ポイントに資本参加、ヤフーも追加出資 (オンライン), 入手先 (<<http://itpro.nikkeibp.co.jp/atcl/news/15/032501071/>>) (参照 2015-03-25).
- [3] Blei, D.M., Ng, A.Y. and Jordan, M.I.: Latent Dirichlet allocation, *Journal of Machine Learning Research*, Vol.3, pp.993–1022 (2003).
- [4] 佐藤一誠: トピックモデルによる統計的潜在意味解析, コロナ社 (2015).
- [5] 岩田具治: 潜在トピックモデルを用いたデータマイニング, *Technical Report of the 1st Workshop on Latent Dynamics* (2010).
- [6] Hofmann, T.: Probabilistic latent semantic indexing, *Proc. 22nd Annual International ACM SIGIR Conference (SIGIR 99)*, pp.50–57 (1999).
- [7] 柳井啓司: 確率トピックモデルによる Web 画像の分類, 第 22 回人工知能学会全国大会 (2008).
- [8] Das, A.S., Datar, M., Garg, A. and Rajaram, S.: Google news personalization: Scalable online collaborative filtering, *Proc. 15th International World Wide Web Conference (WWW'07)*, pp.271–280 (2007).
- [9] 渡邊恵太: トピックモデルと協調フィルタリングに基づくユーザ興味を反映した情報推薦システム, 第 28 回人工知能学会全国大会 (2014).
- [10] Jin, X., Zhou, Y. and Mobasher, B.: Web usage mining based on probabilistic latent semantic analysis, *Proc. 10th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'04)*, pp.197–205 (2004).
- [11] Hofmann, T.: Collaborative filtering via gaussian probabilistic latent semantic analysis, *Proc. 26th Annual*

- International ACM SIGIR Conference (SIGIR'03)*, pp.259–266 (2003).
- [12] Iwata, T. et al.: Sequential Modeling of Topic Dynamics with Multiple Timescales, *ACM Trans. Knowledge Discovery from Data (TKDD)*, Vol.5, No.4, pp.1–19 (2012).
- [13] Iwata, T. and Sawada, H.: Topic Model for Analyzing Purchase Data with Price Information, *Data Mining and Knowledge Discovery*, Vol.26, No.3, pp.559–557 (2013).
- [14] 石垣 司, 竹中 毅, 本村陽一: 百貨店 ID 付き POS データからのカテゴリ別状況依存の変数間関係の自動抽出法, *オペレーションズ・リサーチ*, 経営の科学, Vol.56, No.2, pp.77–83 (2011).
- [15] 小川孔輔: POS とマーケティング戦略, 法政大学産業情報センター (1993).
- [16] 阿部 誠, 近藤文代: マーケティングの科学, 朝倉書店 (2005).
- [17] 青木幸弘, 新倉貴士, 佐々木壮太郎, 松下光司: 消費者行動論, 有斐閣アルマ (2012).
- [18] 守口 剛, 竹村和久: 消費者行動論—購買心理からニューロマーケティングまで, ダイアモンド社 (2012).

付 録

本文中で記載した EC サイトのリストを表 A.1 に示す.

表 A.1 EC サイト名とリンク
Table A.1 site names and links.

EC サイト名	URL
キャス・キッドソン (Cath Kidston)	http://www.cathkidston.jp/
クォーター・リポート	http://www.quarter.co.jp/
ユニクロ	http://www.uniqlo.com/jp/
無印良品	http://www.muji.net/store/
Yahoo!オークション (現 ヤフオク)	http://auctions.yahoo.co.jp/
Yahoo!ショッピング	http://shopping.yahoo.co.jp/
アマゾン (Amazon)	http://www.amazon.co.jp/
伊勢丹	http://isetan.mistore.jp/
シャネル (CHANEL)	http://www.chanel.com/ja-JP/
フォクシー (FOXKEY)	http://www.fokey.co.jp
ヴァンクレフ・アンド・アーペル (Van Cleef & Arpels)	http://www.vancleefarpels.com/
エルメス (HERMES)	http://www.maisonhermes.jp/
ワジャ (waja)	http://www.waja.co.jp/
マーク・バイ・マークジェイコブス (Marc by Marc Jacobs)	http://www.marcjacobs.jp/products/landing.php?category_id=15
ケイト・スペード ニューヨーク (kate spade new york)	http://www.katespade.jp
エルショップ (ELLE SHOP)	http://elleshop.jp/
フラッグショップ (FLAG SHOP)	http://flagshop.jp/
ワールド	http://www.world.co.jp/
三陽商会	http://www.sanyo-shokai.co.jp/
ゾゾタウン (ZOZOTOWN)	http://zozo.jp/
ジーユー (GU)	http://www.gu-japan.com/
フォーエバー 21 (FOREVER21)	http://www.forever21.co.jp/
ジャバリ (Javari)	http://www.amazon.co.jp/gp/feature.html?ie=UTF8&&docId=1000197786
ジル・スチュアート (JILLSTUART)	http://www.jillstuart.jp
マッキントッシュ・フィロソフィー (MACKINTOSH PHILOSOPHY)	http://www.mackintosh-philosophy.com/
ミラベラ (mirabella)	http://flagshop.jp/mirabella/shop/
マガシーク (MAGASEEK)	http://www.magaseek.com/



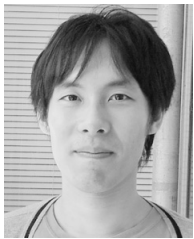
高屋 典子

1987年学習院大学文学部英米文学科卒業。1995年年日本電信電話株式会社 (NTT) 入社。2010年よりNTT サービスエボリューション研究所。現在コ・イノベーションプロジェクトにて新サービスの戦略立案に従事。主幹研究員。日本マーケティングサイエンス学会、日本マーケティング学会会員。



江崎 健司 (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研究員。2010年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等の研究・開発に従事。電子情報通信学会、データベース学会各会員。



石黒 勝彦

2004年東京大学工学部機械情報工学科卒業。2006年同大学大学院情報理工学系研究科知能機械情報学専攻修士課程修了。同年日本電信電話株式会社 (NTT) 入社。現在、NTT コミュニケーション科学基礎研究所研究主任。博士 (工学)。統計的機械学習、時系列データ解析、関係データ解析の研究に従事。電子情報通信学会、IEEE 各会員。



市川 裕介

1994年慶應大学理工学部計測工学科卒業。1996年同大学院修士課程修了。同年日本電信電話株式会社入社。以来、通信履歴活用サービスの研究開発に従事。情報処理学会山下記念研究賞 (2005年) 受賞。