情報処理学会論文誌 Vol.58 No.12 2034–2045 (Dec. 2017)

マルコフ潜在クラスモデルに基づく ECサイトにおける施策実施効果分析に関する一考察

松嵜 祐樹[1,a)] 三川 健太[2,b)] 後藤 正幸[3,c)]

受付日 2017年4月1日, 採録日 2017年9月5日

概要：近年，ECサイト上に蓄積された膨大なユーザ（顧客）の購買行動データを分析し，ユーザごとの特徴を考慮した販売促進施策の重要性が高まっている．これらの施策を実施する際には，施策を実施したからこそ購買するであろうユーザに対してのみの施策実施が重要である．このようなユーザ層を特定するための方法として，ユーザを意味のあるグループに分割するユーザセグメンテーションが考えられる．たとえば，ユーザと購買商品のペアに着目したAspect Modelのような確率モデルによるセグメンテーションは，様々な応用場面において有用性が広く認識されている．一方で，ECサイトには，ユーザと購買商品のペア以外にも，どのページを閲覧したのかという閲覧履歴も存在し，この効果的な活用方法が望まれている．この閲覧履歴の時系列に着目したモデルとして，ページの遷移にマルコフ性を仮定し，潜在クラスを仮定したうえでモデル化したLatent Segment Markov Chain（以下，LSMC）があげられる．しかし，LSMCは閲覧履歴のみに着目したモデルであり，施策の実施やそれによる購買の有無を考慮することができない．そこで本研究では，LSMCを拡張し，閲覧履歴に加え，購買履歴，施策実施有無が考慮可能なモデルを提案し，「施策を実施した場合」と「実施しなかった場合」の購買行動の変化を比較することで施策の実施が効果的なセグメントの特定を行う．

キーワード：顧客セグメンテーション，潜在クラスモデル，マルコフモデル，販売促進，マーケティング

Latent Semantic Markov Model for Effective Promotion Activities in EC Sites

Yuki Matsuzaki[1,a)] Kenta Mikawa[2,b)] Masayuki Goto[3,c)]

Received: April 1, 2017, Accepted: September 5, 2017

Abstract: Recently, it has become popular to purchase product items through E-commerce sites (EC sites), and the internet market scale has been expanding. Under this situation, many EC sites conduct various kinds of sales promotions by analyzing huge amount of customers' purchase histories, and customer segmentation is one of the most important tools in marketing. Particularly, modeling of customers purchase behavior based on probabilistic models such as the Aspect Model (AM) is an attractive way for customer segmentation. The AM focuses on pairs of a customer and an item and it assumes unobserved features such as customers' heterogeneity and items' similarity as latent classes. Although the original AM focuses on pairs of a customer and an item mainly, the data about customers' browsing histories are also available on EC sites. If the model can take in the information of page transitions, it becomes possible to model customers' purchase behavior in detail and make better customer segments. In this paper, we propose a new latent class model that integrates browsing histories in addition to purchase histories by assuming that customers' page transitions can be described by Markov process. An analysis of actual EC site data is demonstrated to clarify the effectiveness.

Keywords: latent class model, markov model, customer segmentation, sales promotion, marketing

[1] 早稲田大学大学院創造理工学研究科経営システム工学専攻
Graduate School of Creative Science and Engineering Waseda University, Shinjuku, Tokyo 169–0072, Japan
[2] 湘南工科大学工学部情報工学科
Department of Information Science, Shonan Institute of Technology, Fujisawa, Kanagawa 251–8511, Japan
[3] 早稲田大学創造理工学部経営システム工学科
School of Creative Science and Engineering Waseda University, Shinjuku, Tokyo 169–0072, Japan
[a)] y_matsuzaki@toki.waseda.jp
[b)] mikawa@info.shonan-it.ac.jp
[c)] masagoto@waseda.jp

1. はじめに

近年，インターネット上のECサイトを通じた商品の購買が頻繁に行われるようになり，その市場規模は引き続き増大傾向にある．このような背景のもと，蓄積された膨大なユーザ（顧客）の購買行動データを分析し，ユーザごとの特徴を考慮したマーケティング施策の重要性が高まっている．これら施策の一例として，本研究で対象とするECサイトで行われているようなリアルタイムに商品の割引クーポンを発行するといった施策があげられる．このような施策においては，商品の購買意思が固まっていたユーザに対する施策実施は売上の減少を導くため望ましくない．したがって，施策を実施したからこそ購買するであろうユーザに対してのみ施策を行うことが重要である [1]．

施策の実施が購買につながるユーザ層を特定するための方法として，ユーザを意味のあるグループに分割するユーザセグメンテーションが考えられる．たとえば，Aspect Model [2]（以下，AM）のような確率モデルによるセグメンテーションは，様々な応用場面において有用性が広く認識されている [3], [4], [5]．AMでは，主にユーザと購買アイテムのペアに対して潜在的なクラスを仮定し，ユーザの嗜好の異質性やアイテムの類似性を考慮した購買行動のモデル化を行っている．

また，ユーザと購買アイテムのペア以外にも，ECサイトのアクセスログデータには，サイト上でどのページを閲覧したのかという閲覧履歴も含まれており，この効果的な活用方法が望まれている．一般に，ユーザの閲覧履歴は，購買履歴よりもデータ量が膨大であり，かつユーザの嗜好や各商品に対する購買意欲の差異が閲覧行動に現れると考えられる．このため，より詳細にユーザの購買行動をモデル化するためには，この閲覧行動をどのように扱うかがポイントとなる．

閲覧履歴の時系列に着目したモデルとして，ページ遷移にマルコフ性を仮定し，潜在クラスを仮定したうえでモデル化したLatent Segment Markov Chain（以下，LSMC）[6] があげられる．LSMCは閲覧ページ遷移に着目したモデル化を行っているが，ECサイトにおける購買有無を考慮したモデルとはなっていない．文献 [7] ではこの点に着目し，LSMCを購買有無が考慮できるように拡張することによって，どのような閲覧行動の後に購買が行われるのかといったより詳細な顧客の購買行動のモデル化を行っている．

ここで，本研究で対象とするクーポン発行施策への援用を考えた場合，クーポン発行による購買行動への効果もユーザごとに異なると考えられる．そのため「施策が実施された場合の購買」と「実施されなかった場合の購買」を区別して施策実施効果をモデル化する必要がある．このようなモデル化ができれば，同一のユーザセグメントについて，「施策を実施した場合」と「実施しなかった場合」の購買行動の変化を比較し，施策の実施が効果的なセグメントを特定することができ，的確なターゲティングにつながると考えられる．

そこで本研究では，ユーザの閲覧履歴に加えて，購買有無を考慮したモデルおよび，購買有無と施策実施の効果を考慮したモデルの構築を行う．モデル化に際しては，文献 [7] のモデルを拡張することで施策の実施が効果的なユーザ層の特定を可能とする新たなモデルを構築する．また，大手総合通販カタログサイトのアクセスログデータおよび施策実施履歴のデータに対して，提案モデルを適用することでモデルの有用性を示す．

2. 準備

2.1 対象問題とデータ概要

本研究では，ECサイトのアクセスログ解析を行う企業が保有するデータ（大手総合通販カタログサイトのアクセスログデータ）を分析対象とする．以下では，ユーザがサイトを訪問してから離脱するまでの一度の閲覧行動をセッションと呼ぶこととし，これに一意なIDが付与されているものとする．さらに，このIDに閲覧履歴と購買有無が紐付けられている．ECサイトの各ページにはそれぞれ “item” や “category” といったような12種類のページタイプが付与されており，本研究ではこのページタイプの閲覧遷移を閲覧履歴として使用する．**表 1** に本研究で対象としたECサイト上で定義されているページタイプ名とその概要を示す．本研究では，分析対象データを取り扱っている企業において1セッションを1ケースとしている点，ならびにユーザの購買行動は毎回の訪問時に異なる可能性があるという前提のもと，セッション単位での分析を検討するものとする．なお，実際には，商品特性によってはユーザが検索や購入においてセッションをまたぐことも考えられるが，本研究で対象としているECサイトの取扱商品は長期間の検討がなされるカテゴリではないため，このようなケースは考慮外とした[*1]．本研究で提案するモデルでは，同一ユーザが購買時の目的やシチュエーションなどによって異なる閲覧行動をとったり，異なる嗜好で購買の判断をするという状況もまた表現することが可能となっている．

対象とするECサイトでは，すでにある種の機械学習モデルを用いてリアルタイムのクーポン発券を実施している

[*1] 各個々のユーザが識別できる場合には，1つのセッションを1ケースとして取り扱うのではなく，1人のユーザの一連の閲覧行動を1ケースとして扱うことも可能である．その場合，1人のユーザがある程度の期間にわたって商品の購入を迷い，セッションをまたぐことも考えられるため，1セッションを1ケースとして扱う状況とは異なる結果が得られる可能性がある．しかし，ECサイトでログインせずに閲覧や検索をするユーザもいることから，個々のユーザを識別することが難しいケースも多い．このような状況を考慮し，本研究では広く導入が容易と考えられるセッション単位のモデル化を行う．

表 1 対象 EC サイトにおけるページタイプの種類およびその概要
Table 1 The page types and its roles in the EC site.

ページタイプ名	概要
top	対象サイトのトップページ
search	検索結果が表示されているページ
category	同一カテゴリの商品が一覧表示されるページ
item	商品の詳細ページ
cart	カートページ
cart_form	購入前の必要事項記入ページ
conversion	「ご購入ありがとうございました」と表示されるページ
registration_form	会員登録のための必要事項記載ページ
registration	会員登録に関連するページ
login	ログインのための ID およびパスワードの入力ページ
sale	セール商品が表示されるページ
catalog	カタログに記載された商品が表示されるページ

が[*2]，その施策効果の検証を行うために，サイトを訪れたセッションに対してランダムに A，B のラベルをつけ，「施策を行うセッション（ラベル A)」と「施策を行わないセッション（ラベル B)」の購買割合を比較する A/B テストを行っている[*3]．なお，ユーザのプライバシー保護の観点から，デモグラフィックデータは分析対象としない．

本研究で対象としたセッションデータの基本統計量については付録参照のこと．

2.2 閲覧履歴の表記

本研究で対象とする閲覧履歴について説明を行う．いま，I 個からなるセッション集合を $\mathcal{S} = \{s_i : 1 \leq i \leq I\}$，$J$ 種類からなるページタイプ集合を $\mathcal{K} = \{k_j : 1 \leq j \leq J\}$ とする．i 番目のセッション s_i で閲覧されたページタイプの数を T_i とすれば，閲覧履歴系列 $\boldsymbol{x}_i$ は以下で与えられる．

$$\boldsymbol{x}_i = x_0^i x_1^i \cdots x_{T_i-1}^i \tag{1}$$

ただし，x_t^i は i 番目のセッション s_i で t 番目に閲覧されたページタイプで，$x_t^i \in \mathcal{K}$ を満たす．また，J 種類からなるページタイプは，対象とする EC サイトで定義されているものを用いる．

2.3 関連研究

2.3.1 潜在クラスモデル

マーケティングなどの分野においては，消費者の嗜好や行動は均一的ではなく，異なる特徴を有したグループが混在していることを仮定することが多い．これは，ユーザの嗜好の多様性を表現するための合理的な考え方の 1 つであり，マーケティング分野ではカスタマーセグメンテーション（市場細分化）といった考え方に通じている [8]．このような対象に対して適用され，有用性が認められる統計モデルが潜在クラスモデルである [9], [10], [11]．潜在クラスモデルは，観測される統計データの背後に観測できないクラスという離散変数の存在を仮定し，観測データから EM アルゴリズム [12], [13] などを用いてパラメータの推定を行う．

マーケティング分野における潜在クラスモデルについては比較的古くから議論がなされており，市場細分化との関連について論じた Green らの研究 [14] や製薬分野の市場細分化に適用した Bassi の研究 [4]，クラスタリングモデルとして k-means 法との比較 [15] などがある．比較的最近の大規模データに対して適用した研究としては，Hofmann による Aspect Model [16] がよく知られている．Hofmann のモデルは PLSA（Probabilistic Latent Semantic Analysis）[2] モデルとも呼ばれ，もともとはテキストデータ分析のためのトピックモデルの 1 つとして提案された．その後，協調フィルタリングによるレコメンデーション [16] に適用され，その有効性が示されたことから，様々な拡張がなされている [17], [18], [19]．これらの潜在クラスモデルはユーザの購買履歴データの分析にも有用であり，そのための分析モデルとしても様々なタイプが提案されている [20], [21], [22]．

本研究では，これらの従来の潜在クラスモデルに対し，EC サイト上のページ遷移というユーザのページ閲覧行動に対する潜在クラスモデルを考える．このようなユーザのページ閲覧行動に対して潜在クラスモデルを用いてモデル化した研究としては，LSMC（Latent Segment Markov Chain）[6] がある．本研究では，この LSMC をベースとして，クーポン発券などのマーケティング施策に対するユーザの購買行動を分析可能とするモデルに拡張を行う．従来研究としての LSMC については，次項で概要を示す．

2.3.2 Latent Segment Markov Chain

LSMC はユーザの閲覧行動にマルコフ性 [9] を仮定し，

[*2] 現在導入されているクーポン発券ロジックについては詳細を示すことができないが，本研究で導入する潜在クラスモデルのように，対象の異質性を考慮した方法ではなく，後で示すように本研究の提案手法と組合せて用いることが可能である．
[*3] 本研究では，この A/B テストで付与されたラベルデータも評価用に用いる．

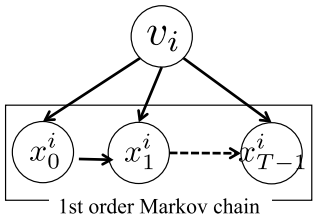


図 1 LSMC のグラフィカルモデル [6]
Fig. 1 The graphical of LSMC.

潜在クラスモデルとして表現した統計モデルである．POS データのような，あるユーザの購買した 1 つの商品といったようなユーザと商品のペアではなく，ユーザの購買履歴のような複数の商品を含むデータに対するモデル化としては，ユーザの 1 トランザクションにおける複数アイテムの購買を表現したモデル [20], [23] があげられる．しかしながら，このようなモデルでは，対象としているデータの時系列を考慮しておらず，本研究で対象とするようなあるページを閲覧したのちに次のページを閲覧するといった状況をモデルとして考慮できない．そこで，LSMC では，t 番目の閲覧ページタイプは，$t-1$ 番目の閲覧ページタイプのみに依存するという，1 次のマルコフ性を仮定することでセッション s_i の持つ閲覧履歴系列 $\boldsymbol{x}_i$ の確率構造を潜在クラスを条件とする条件付き 1 次マルコフモデルとしてモデル化している．このようなモデル化を行うことによって，トップページからセールページに遷移しやすいユーザやカートページに遷移しにくいユーザなど閲覧行動の類似性や異質性を考慮したユーザセグメントのモデル化が可能となる．LSMC のグラフィカルモデルは**図 1** のようになる．

いま，L 個からなる潜在クラス集合を $\mathcal{Z} = \{z_l : 1 \leq l \leq L\}$ と定義すれば，ある潜在クラス v_i に所属するセッション s_i における閲覧履歴 $\boldsymbol{x}_i$ の生起確率は以下の式 (2) のように表現される．ただし，s_i が所属する潜在クラス v_i は，$v_i \in \mathcal{Z}$ を満たす．

$$P(\boldsymbol{x}_i|v_i) = P(x_0^i|v_i) \prod_{t=1}^{T_i-1} P(x_t^i|x_{t-1}^i, v_i) \tag{2}$$

なお，式 (2) において，$P(x_0^i|v_i)$ は初期分布であり，$P(x_t^i|x_{t-1}^i, v_i)$ は s_i において $t-1$ 番目にページタイプ $x_{t-1}^i \in \mathcal{K}$ を閲覧したのちに，t 番目にページタイプ $x_t^i \in \mathcal{K}$ を閲覧する確率である．

2.3.3 閲覧履歴および購買行動を考慮したマルコフ潜在クラスモデル [7]

本節では，閲覧履歴に加えて，購買の有無を考慮できるよう LSMC を拡張したモデル（以下，Matsuzaki らのモデル）[7] について述べる．このような拡張を行うことによって，EC サイトにおける購買行動をより詳細に記述することが可能となり，購買につながる閲覧行動を特定することができると考えられる．

Matsuzaki らのモデルでは，LSMC で考慮している閲覧履歴に加えて，購買の有無を考慮できるように拡張を行う．LSMC で定義されている変数と集合に加えて，i 番目のセッション s_i における購買の有無を表す変数 w_i を以下のように定義する．なお，式 (3) 中の y は「購買が起きる」という事象，$\bar{y}$ はその余事象である．

$$w_i = \begin{cases} 1, & \text{事象 } y \text{ が生起するとき} \\ & (s_i \text{ で購買が起こる場合}) \\ 0, & \text{事象 } y \text{ が生起しないとき} \\ & (s_i \text{ で購買が起こらない場合}) \end{cases} \tag{3}$$

ここで，s_i の閲覧履歴 $\boldsymbol{x}_i$ と購買有無 w_i に対応する潜在変数を v_i とすれば，i 番目の完全データは $(\boldsymbol{x}_i, w_i, v_i)$ と表される．なお，$v_i \in \mathcal{Z} = \{z_l : 1 \leq l \leq L\}$ を満たす．よって，i 番目の完全データ $(\boldsymbol{x}_i, w_i, v_i)$ についての確率モデルは以下のように表される．

$$P(\boldsymbol{x}_i, w_i, v_i) = P(v_i)P(\boldsymbol{x}_i|v_i)P(w_i|v_i) \tag{4}$$

ただし，$P(w_i|v_i) = P(y|v_i)^{w_i} P(\bar{y}|v_i)^{1-w_i}$ とし，$P(y|z)$ はある潜在クラス $z \in \mathcal{Z}$ において購買が起きるか否かという 2 値の事象に対してベルヌーイ分布を仮定している．

よって，潜在変数 $v_i = z_l$ の下での閲覧履歴 $\boldsymbol{x}_i$，購買有無 w_i の生起確率 $P(\boldsymbol{x}_i, w_i|v_i = z_l)$ は以下のようになる．

$$\begin{aligned} &P(\boldsymbol{x}_i, w_i|v_i = z_l) \\ &= P(\boldsymbol{x}_i|z_l)P(y|z_l)^{w_i} P(\bar{y}|z_l)^{1-w_i} \\ &= P(x_0^i|z_l)P(y|z_l)^{w_i} P(\bar{y}|z_l)^{1-w_i} \prod_{t=1}^{T_i-1} P(x_t^i|x_{t-1}^i, z_l) \\ &= \prod_{j=1}^{N} \lambda_{lj}^{\delta(x_0^i=k_j)} \prod_{j=1}^{N}\prod_{m=1}^{K} a_{ljm}^{n_{ijm}} \gamma_l^{w_i} (1-\gamma_l)^{1-w_i} \end{aligned} \tag{5}$$

ただし，$\lambda_{lj} = P(x_0^i = k_j|v_i = z_l)$ とし，$\delta(x_0^i = k_j)$ は $x_0^i = k_j$ のとき 1 となるインジケータ関数，$a_{ljm} = P(x_t = k_j|x_{t-1} = k_m, z_l)$，$n_{ijm}$ はページタイプ k_m から k_j への遷移回数，$\gamma_l = P(y|z_l) = P(y|v_i = z_l)$ である．なお，このモデルは観測されない変数である潜在変数を含むため，EM アルゴリズム [12], [13] を用いてパラメータの推定を行う．

従来手法の LSMC をはじめ，Matsuzaki らの手法，ならにび本研究で提案するモデルでは，潜在クラスが与えられたもとでの条件付独立の構造を仮定することで，様々な嗜好を持った顧客が混在する場合の閲覧履歴と購買行動の従属関係をモデル化している．潜在クラスモデルを仮定することで，閲覧履歴と購買行動の関係性が異なるグループが混在する対象をモデル化することが可能となる．

3. 閲覧履歴，購買有無および施策実施有無を考慮したマルコフ潜在クラスモデル

3.1 概要

Matsuzaki らのモデルを用いて施策を実施すべきクラス

を特定する場合には，モデル学習後にそれぞれの潜在クラスについて施策実施の効果（以下，施策効果）を算出し，その値が高い潜在クラスを施策実施すべきクラスと判断すればよい．一方，学習データ中には，現行のクーポン発行ロジックによって「施策が実施されたセッション」と「実施されなかったセッション」が含まれている．本研究で対象とする事例では，すでにリアルタイムに割引クーポンが発行されており（4.1，4.2 節参照），本研究では，その効果をさらに高めるような対象セグメントの特定のために，これらの現行のクーポン発券ロジックによる施策実施の有無データを活用する．すなわち，施策を実施されたセッションと実施されなかったセッションでの購買行動の変化によって潜在クラスごとの施策効果を算出することができる．

しかしながら，Matsuzaki らのモデルを用いて施策実施の意思決定を行う場合には，モデルの学習後に各潜在クラスに含まれる「施策が実施されたセッション」と「実施されなかったセッション」の購買割合を用いて施策効果を算出することになるため，施策効果をモデルから間接的に算出していることになる．つまり，このモデルは学習時に施策実施有無と購買の因果関係を表現していないため，直接的な施策効果をモデルとして考慮できていないといえる．この施策効果についても，施策実施が効果的なクラスやあまり効果がないクラスなど潜在クラスによって差異があると考えられる．この施策効果の異質性を考慮するためには，「施策が実施された場合の購買」と「実施されなかった場合の購買」を区別して施策実施と購買の因果関係をモデル化する必要がある．

そこで本節では，前項の Matsuzaki らのモデルを拡張し，施策実施の有無を直接的に考慮したモデル（以下，提案モデル）の提案を行う．このように，対象 EC サイトにおいて施策実施されたセッションに注目し，施策実施有無を考慮したモデル化を行うことで，「施策を実施した場合」と「実施しなかった場合」の購買確率が推定可能となり，施策実施が効果的な潜在クラスを特定することができると考えられる．提案モデルのグラフィカルモデルは**図 2** のように表される．なお，図 2 中の観測変数 d_i は，i 番目のセッション s_i における施策実施有無を表す変数であり，式 (6) のように定義される．式 (6) における c は「施策が実施される」という事象であり，$\bar{c}$ はその余事象である．

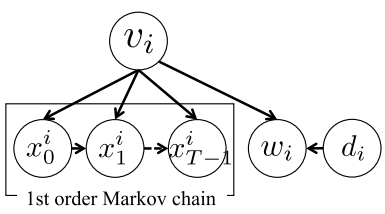


図 2 提案モデルのグラフィカルモデル
Fig. 2 The graphical of the proposed model.

$$d_i = \begin{cases} 1, & \text{事象 } c \text{ が生起するとき} \\ & (s_i \text{ で施策が実施された場合}) \\ 0, & \text{事象 } c \text{ が生起しないとき} \\ & (s_i \text{ で施策が実施されなかった場合}) \end{cases} \tag{6}$$

3.2 定式化

提案モデルでは，Matsuzaki らのモデルを拡張し，施策実施の有無を条件とした購買を考慮できるように拡張を行う．Matsuzaki らのモデルで定義されている集合と変数に加えて，i 番目のセッション s_i について，施策実施の有無を表す事象 $c, \bar{c}$ に対応する観測変数 d_i を導入する．ここで，s_i の閲覧履歴を $\boldsymbol{x}_i$，購買有無 w_i に対応する潜在変数を v_i とすれば，i 番目の完全データは $(\boldsymbol{x}_i, w_i, d_i, v_i)$ と表される．ただし，$v_i \in \mathcal{Z} = \{z_l : 1 \le l \le L\}$ を満たす．よって，i 番目の完全データの確率モデルは以下のように表される．

$$P(\boldsymbol{x}_i, w_i, v_i | d_i) = P(v_i) P(\boldsymbol{x}_i | v_i) P(w_i | v_i, d_i) \tag{7}$$

なお，$P(w_i | v_i, d_i)$ は

$$\begin{aligned} P(w_i | v_i, d_i) &= \{P(y|c, v_i)\}^{d_i w_i} \{P(\bar{y}|c, v_i)\}^{d_i(1-w_i)} \\ &\cdot \{P(y|\bar{c}, v_i)\}^{(1-d_i)w_i} \{P(\bar{y}|\bar{c}, v_i)\}^{(1-d_i)(1-w_i)} \end{aligned} \tag{8}$$

を示すものとし，$P(y|c, z)$ はある潜在クラス $z \in \mathcal{Z}$ に所属するセッションに施策が実施されたという条件のもとで，また，$P(y|\bar{c}, z)$ はされないという条件のもとで購買が起こる確率であり，それぞれ購買が起きるか否かという 2 値の事象に対してベルヌーイ分布を仮定している．また，潜在変数 $v_i = z_l$ および施策実施有無を表す変数 d_i が与えられたもとでの閲覧履歴 $\boldsymbol{x}_i$，購買有無 w_i の生起確率 $P(\boldsymbol{x}_i, w_i | v_i = z_l, d_i)$ は以下のようになる．

$$\begin{aligned} &P(\boldsymbol{x}_i, w_i | v_i = z_l, d_i) \\ &= P(\boldsymbol{x}_i | z_l) \{P(y|c, z_l)\}^{d_i w_i} \{P(\bar{y}|c, z_l)\}^{d_i(1-w_i)} \\ &\quad \cdot \{P(y|\bar{c}, z_l)\}^{(1-d_i)w_i} \{P(\bar{y}|\bar{c}, z_l)\}^{(1-d_i)(1-w_i)} \\ &= \prod_{j=1}^{J} \lambda_{jl}^{\delta(x_0^i = k_j)} \prod_{j=1}^{J} \prod_{m=1}^{J} a_{jml}^{n_{ijm}} {\gamma_l^1}^{d_i w_i} {\overline{\gamma}_l^1}^{d_i(1-w_i)} \\ &\quad \cdot {\gamma_l^0}^{(1-d_i)w_i} {\overline{\gamma}_l^0}^{(1-d_i)(1-w_i)} \end{aligned} \tag{9}$$

ただし，$\gamma_l^1 = P(y|v_i = z_l, d_i = 1)$，$\gamma_l^0 = P(y|z_l, d_i = 0) = P(y|v_i = z_l, d_i = 0)$，$\overline{\gamma}_l^1 = 1 - \gamma_l^1$，$\overline{\gamma}_l^0 = 1 - \gamma_l^0$ である．

3.3 パラメータ推定

提案モデルについても潜在変数を含むため，EM アルゴリズム [12], [13] を用いてパラメータの推定を行う．以下に E-step, M-step それぞれの更新式を示す．

E-step

$$P(z_l|\boldsymbol{x}_i, w_i, d_i) = \frac{\pi_l P(\boldsymbol{x}_i, w_i|z_l, d_i)}{\sum_{l'=1}^{L} \pi_{l'} P(\boldsymbol{x}_i, w_i|z_{l'}, d_i)} = \alpha_{il} \quad (10)$$

M-step

$$\pi_l = \frac{1}{I}\sum_{i=1}^{I} \alpha_{il} \quad (11)$$

$$\lambda_{jl} = \frac{\sum_{i=1}^{I} \alpha_{il}\delta(x_0^i = k_j)}{\sum_{i=1}^{I} \alpha_{il}} \quad (12)$$

$$a_{jml} = \frac{\sum_{i=1}^{I} \alpha_{il} n_{ijm}}{\sum_{m=1}^{J}\sum_{i=1}^{I} \alpha_{il} n_{ijm}} \quad (13)$$

$$\gamma_l^1 = \frac{\sum_{i=1}^{I} \alpha_{il} d_i w_i}{\sum_{i=1}^{I} \alpha_{il} d_i} \quad (14)$$

$$\gamma_l^0 = \frac{\sum_{i=1}^{I} \alpha_{il} (1-d_i) w_i}{\sum_{i=1}^{I} \alpha_{il} (1-d_i)} \quad (15)$$

EM アルゴリズムでは，式 (16) で表される完全データの対数尤度が収束するまでパラメータの更新を行う．ただし，$\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_I)$，$\boldsymbol{W} = (w_1, \ldots, w_I)$，$\boldsymbol{D} = (d_1, \ldots, d_I)$，$\boldsymbol{V} = (v_1, \ldots, v_I)$ とする．

$$\log P(\boldsymbol{X}, \boldsymbol{W}, \boldsymbol{D}, \boldsymbol{V}) = \sum_{i=1}^{I} \log P(\boldsymbol{x}_i, w_i, d_i, v_i) \quad (16)$$

4. 実データ分析への適用

4.1 概要

提案モデルの有用性を検証するため，大手総合カタログ通販サイトにおける購買履歴，閲覧履歴および施策実施有無データを用いた実験を行った．対象 EC サイトの各ページにはそれぞれ “item” や “category” といったような 12 種類のページタイプ（表 1 参照）が付与されており，本実験ではこのページタイプの閲覧遷移を閲覧履歴として使用する．また，対象 EC サイトでは，すでにリアルタイムに割引クーポンを発行するという施策を実施しており，このクーポン発行の有無を施策の実施有無データとしてモデルの学習に用いる．

4.2 モデルの評価方法

対象 EC サイトでは，リアルタイム割引クーポン発行施策の効果を検証するために，マーケティングなどの分野で広く用いられている A/B テストが行われている．この A/B テストでは，サイトを訪れたすべてのセッションに対してランダムに**表 2** のような A，B のラベルをつけ，現在使われているクーポン発券ロジックによりクーポン発行すべきと判断されたセッション（クーポン対象セッション）のうち，A 群であるセッションのみにクーポン発行が行われ，B 群のセッションには発行が行われない．A，B 群の差を見ることにより，クーポン発行の効果を測定することが可能となる．

表 2 データのラベル付け
Table 2 Labels for using data.

		A/B テストのラベル	
		A	B
クーポン対象	対象	A_c	B_c
	非対象	A_{nc}	B_{nc}

このラベルを用い，モデル学習により推定された A_c 群の購買確率と B_c 群の購買確率の比（$P(y|z,c)/P(y|z,\bar{c})$）によって各潜在クラスの施策効果を推定する．この値が高い潜在クラスを提案モデルによって特定される「クーポン発行効果が高いクラス」とする．また，各潜在クラスに所属するセッションのうち A_c，B_c 群になっているセッションの割合を見ることによって，実験により施策効果が高いと判断されたクラスに所属するセッションに対し，現行の施策がどれだけクーポン対象とできているのかを検証する．これによって，対象 EC サイトで行われたクーポン発行施策の評価を行う．

4.3 実験条件

実験には，2016 年 4 月 1 日から 30 日までの 1 カ月間で蓄積された閲覧，購買履歴および施策実施有無データを用いる．施策実施有無をモデルとして考慮した場合，クーポン対象セッション（ラベル A_c，B_c）の割合が小さいため，学習データ中のクーポン発行セッションの割合が少なくなってしまい，全セッションを用いるとモデル学習時にクーポン対象セッションが相対的に軽視されてしまう．したがって，提案モデルでは，クーポン対象セッション（ラベル A_c，B_c）により注目するために，「クーポン対象セッション数」と「非クーポン対象セッション数」が同程度になるようにデータをランダムにサンプリングし，そのデータを学習データとしてモデルの学習を行う．学習データの総セッション数は 109,708 件，総閲覧件数は 2,222,716 件である．

なお，閲覧されるページの性質を考慮し，本実験では閲覧端末が PC であるデータのみを用いるものとした．また，潜在クラスの数については，AIC や BIC などの情報量基準を適用する方法も考えられるが，本研究で対象とするモデルでは，最尤推定量が漸近的に正規分布に従うことを保証できず，理論的な保証が得られない．潜在クラス数を大きくしすぎるとクラスあたりのデータ数が減少し，パラメータの推定精度が下がるため，適切な潜在クラス数を推定精度や予測精度の観点から探索するという方法も考えられる．しかしながら，単に結果が改善されるというだけでなく，そのモデルの内容を適切に解釈でき，なぜ良くなる

表 3 実験結果
Table 3 The result of the experiment.

潜在クラス	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}	全体
所属セッション数	9,173	6,065	8,402	9,922	18,951	17,226	5,076	8,344	7,044	19,506	109,708
混合比	0.08	0.06	0.08	0.09	0.17	0.16	0.05	0.08	0.06	0.18	1.00
A_c 購買確率	0.394	0.171	0.088	0.023	0.045	0.263	0.223	0.156	0.084	0.047	0.144
B_c 購買確率	0.309	0.099	0.060	0.016	0.040	0.225	0.152	0.140	0.065	0.034	0.113
施策実施効果 (A_c/B_c)	1.28	1.73	1.46	1.42	1.12	1.17	1.46	1.11	1.28	1.40	1.28
カバー率	48%	35%	45%	57%	58%	59%	55%	57%	52%	33%	50%

のかが説明できることもまた実応用においては重要な点となる．このような理由より，各潜在クラスの所属セッション割合がある程度のボリュームを持ちつつ，クーポン効果と閲覧傾向の差異について解釈可能であるような潜在クラス数を事前実験により探索するものとした．この事前実験では潜在クラスの数を 2 から 22 まで 1 刻みに変化させ，加えて 40，80，100 と設定したもとで実験を行い，その中で特徴的な結果を得られた潜在クラス数，すなわち $L = 10$ と設定するものとした．

EM アルゴリズムの収束値は一般に初期値に依存してしまうという問題があるため，本実験では 3 回の異なる初期値による実験を行っている．これらによって得られた潜在クラス全体を確認したところ，ほぼ同様の結果となった．このため，本論文ではその中の 1 つの実験結果について示すものとする．

4.4 実験結果・考察

A_c 群の購買確率 $P(y|z,c)$ と B_c 群の購買確率 $P(y|z,\bar{c})$ の値や施策実施効果，カバー率，混合比[*4]に着目し，分析を行った．**表 3** に 10 個の潜在クラスそれぞれの結果を示す．ここでは，これらの値を総合的に勘案し，特徴的な結果を示した 3 つの潜在クラス（z_3，z_5，z_6）について考察を行う．なお，施策実施効果は，A_c 群と B_c 群の購買確率の比（$P(y|z,c)/P(y|z,\bar{c})$）により定義し，この値が高いほど施策効果が高いと解釈できる．また，カバー率は，各潜在クラスに所属するセッションのうちクーポン対象（ラベル A_c，B_c）となったセッションの割合と定義する．

表 3 より，z_6 について A_c，B_c の購買確率が高いことが分かる．しかし，施策実施効果は 1.17 と全体の施策実施効果と比較して低い値を示している．すなわち，z_6 には，クーポン発行を行わずとも購買に至るセッションが多く集まっていると考えられる．したがって，この潜在クラスへのクーポン発行は，全体と比較して余分な割引の実施につながる可能性がある，すなわち機会損失が発生してしまう可能性があり，クーポン発券の効果は相対的に低いと考えられる．このため，実際の適用場面においては，クーポン発券を積極的に行うべき潜在クラスではないと考えられる．一方 z_3 は，B_c の購買確率が低い値をとっているが，施策実施効果が 1.46 と高い値を示している．このことから，z_3 に所属するセッションは購買に対する意欲があまり高くないが，クーポン発行によって購買が促されるといえるため，重点的に施策を実施すべきである．また，z_5 については，B_c の購買確率，施策実施効果ともに低い値をとっている．つまり，z_5 に所属するセッションはそもそも購買をする意欲の低い閲覧行動であると考えられるため，施策を実施すべきでないといえる．

次にカバー率を見ることによって，現行のクーポン発行アルゴリズムがそれぞれの潜在クラスに所属するセッションをどの程度クーポン対象として検出しているのかについて分析を行う．z_6 は購買確率が高く，クーポンを発行せずとも購買を行うクラスであるが，そのカバー率は 59%と高い値をとっている．一方で施策実施効果が高い値を示していた z_3 のカバー率は 45%と比較的低い値となっていることが分かる．このことから，現行のクーポン発行アルゴリズムは，施策実施効果が低く，そもそもの購買確率が高いセッションをクーポン対象として検出しがちであると考えられる．すなわち，施策実施効果の高い z_3 に所属するセッションに対して，積極的にクーポン発行を行うことで，施策効果の向上が期待できるといえる．また，購買意欲が低いと考えられる z_5 のカバー率は 58%と高い値をとっていることから，現状ではクーポン発行が効果的でないセッションに対するクーポン発行が行われる傾向にあることが分かる．このクラスへのクーポン発行を控えることによってクーポン対象セッションの質を向上させることができると考えられる．

4.5 詳細分析

前節の分析から，潜在クラス z_3 は購買確率は比較的低いが施策実施効果が高いクラスであることが明らかとなった．本節では，この潜在クラス z_3 に着目し，初期ページ

*4 混合比は，各潜在クラスのもとでの確率構造をどのような比率で混合しているかを決めるモデルのパラメータと定義する．

z_3	top	search	category	item	cart	cart_form	conversion	registration_form	registration	login	sale	catalog		landing page
top	8%	3%	75%	5%	3%	0%	0%	0%	0%	3%	2%	1%		35%
search	4%	51%	19%	22%	1%	0%	0%	0%	0%	0%	2%	0%		2%
category	1%	1%	64%	32%	1%	0%	0%	0%	0%	0%	0%	0%		44%
item	1%	1%	67%	5%	25%	0%	0%	0%	0%	0%	1%	0%		11%
cart	5%	0%	10%	48%	20%	5%	0%	0%	0%	11%	0%	0%		4%
cart_form	2%	0%	1%	2%	3%	66%	26%	0%	0%	1%	0%	0%		0%
conversion	64%	0%	13%	10%	1%	0%	1%	0%	0%	8%	2%	0%		0%
registration_form	1%	0%	1%	0%	1%	10%	0%	80%	2%	6%	0%	0%		0%
registration	61%	0%	11%	0%	10%	4%	0%	9%	6%	0%	0%	0%		0%
login	4%	0%	3%	2%	4%	38%	0%	7%	0%	41%	0%	0%		0%
sale	3%	3%	14%	7%	1%	0%	0%	0%	0%	0%	72%	0%		2%
catalog	15%	1%	8%	3%	42%	0%	0%	0%	0%	0%	0%	30%		0%

図 3 初期ページの確率分布およびページ遷移確率分布（z_3）

Fig. 3 The probability distributions of the initial page and page transitions (z_3).

z_6	top	search	category	item	cart	cart_form	conversion	registration_form	registration	login	sale	catalog		landing page
top	13%	3%	30%	7%	2%	0%	0%	0%	0%	2%	42%	0%		17%
search	5%	56%	13%	12%	1%	0%	0%	0%	0%	0%	13%	0%		2%
category	7%	3%	60%	16%	1%	0%	0%	0%	0%	0%	13%	0%		11%
item	1%	0%	3%	4%	8%	0%	0%	0%	0%	0%	84%	0%		19%
cart	3%	0%	1%	50%	25%	3%	0%	0%	0%	6%	12%	0%		1%
cart_form	3%	0%	1%	3%	4%	63%	25%	0%	0%	1%	0%	0%		0%
conversion	74%	0%	4%	7%	0%	0%	0%	0%	0%	6%	9%	1%		0%
registration_form	0%	0%	1%	0%	2%	8%	0%	79%	3%	7%	0%	0%		0%
registration	31%	0%	0%	0%	44%	0%	0%	24%	0%	0%	0%	0%		0%
login	6%	0%	2%	5%	4%	32%	0%	3%	0%	40%	7%	0%		0%
sale	1%	0%	1%	49%	1%	0%	0%	0%	0%	0%	49%	0%		50%
catalog	17%	1%	3%	3%	41%	0%	0%	0%	0%	2%	0%	33%		0%

図 4 初期ページの確率分布およびページ遷移確率分布（z_6）

Fig. 4 The probability distributions of the initial page and page transitions (z_6).

(landing page) の確率分布やページ遷移確率の分布の分析を行うことにより，どのような特徴を持つユーザセグメントに対する施策実施が効果的であるのかについて考察を行う．具体的には，提案モデルの学習によって得られるパラメータであるそれぞれの潜在クラスの初期ページ（landing page）の確率分布とページ遷移確率の分布の比較を行う．1 つの潜在クラスにはマルコフモデルにおける 1 つの状態が対応しており，その状態における前述した 2 種類の確率分布を比較することによって，それぞれの潜在クラスについて詳細な分析を行う．以下の**図 3**，**図 4** に潜在クラス z_3, z_6 の初期ページの確率分布やページ遷移確率の分布を示す．なお，比較のために購買確率が高いが，施策実施効果が低いクラスである z_6 の初期ページの確率分布やページ遷移確率の分布も図 4 に示した．

図 3，4 の 12×12 行列で表される部分は，パラメータ a_{ljm} の値であり，潜在クラス z_l におけるページタイプ k_m（行）から k_j（列）への遷移確率を示したものである．また，表右部の “landing page” は初期ページの確率分布であり，表の下部にある出現確率は，以下の式 (17) により算出され，それぞれのページがその潜在クラスにおいてどの程度出現しやすいかを表す．

$$P(k_j|z_l) = \frac{\sum_{m=1}^{J} P(k_j|k_m, z_l)}{\sum_{j=1}^{J}\sum_{m=1}^{J} P(k_j|k_m, z_l)} \tag{17}$$

なお，図中では，値が高いほど色が濃くなるようにセルが着色されている．

まず，図 3 における landing page 列に注目してみると，top，category が高い値を示している．このことから，この潜在クラスに所属しているユーザは当該サイトの閲覧，購買を目的としてアクセスしている可能性が高い．また，全体的な閲覧の傾向を比較するために，z_3 とそれ以外の潜在クラスにおけるページ遷移確率の分布の比較を行う．具体的には，着目する潜在クラス z_l と比較対象となる潜在クラス z について，j 番目のページタイプ k_j から J 種類あ

表 4 z_3 とその他潜在クラスとのページ遷移確率の近さ（$KLD(z_3, z)$）
Table 4 Distance between the transition probability of z_3 and others ($KLD(z_3, z)$).

潜在クラス	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}
KLD	6.24	2.23	0	6.81	36.13	7.63	64.97	5.02	19.31	16.23
購買確率	0.394	0.171	0.088	0.023	0.045	0.263	0.223	0.156	0.084	0.047

るページへの遷移確率の分布の近さを KL 情報量によって評価し，その和によって全体のページ遷移確率の比較を行う．着目する潜在クラス z_l と比較対象となる潜在クラス z のページ遷移確率の近さ（$KLD(z_l, z)$）は以下の式 (18) によって算出する．

$$KLD(z_l, z) = \sum_{j=1}^{J} \sum_{m=1}^{J} P(k_m|k_j, z_l) \log \frac{P(k_m|k_j, z_l)}{P(k_m|k_j, z)} \tag{18}$$

式 (18) を用いて z_3 とその他の潜在クラスとのページ遷移確率の近さを算出したものが**表 4** である．

表 4 より，購買確率が高い潜在クラスを見てみると，z_1, z_2, z_6, z_8 と，z_3 との KLD の値が比較的小さいものとなっていることが分かる．このことから，z_3 に所属するセッションは，購買確率が高い潜在クラスと類似する閲覧傾向を持っているが，購買には至りにくいという特徴があることが明らかとなった．

次に，購買確率が高く施策効果が低い z_6 との比較を行う．購買に直結するページ遷移に着目すると，z_3 の item から cart への遷移確率が 25%であるのに対し，z_6 は 8%と低い値をとっていることから，z_6 に所属するセッションでは，商品をカートに入れる行動に至りにくいことが分かる．また，z_3 の cart から login への遷移確率が 11%であるのに対し，z_6 は 6%と低い値を示していることから，z_6 に所属するセッションでは，商品をカートに入れたあとに，ログインして商品の購買に必要な行動をとりにくいことも分かる．また，図 4 における landing page に着目すると，sale ページの値が高いことが分かる．このことから，この潜在クラスに所属しているユーザは何らかのセール情報により当該サイトを訪れている可能性があり，事前の購入意思がそれほど高くないということも考えられる．したがって，上記の潜在クラス z_3, z_6 の特徴から z_6 に所属するセッションは，「購買確率の高い z_3 と類似した閲覧行動をとるが，購買に至らないセッション」であるといえる．

また，sale ページへの遷移確率に着目すると，top から sale，item から sale，category から sale などの遷移確率において z_6 が高い値を示していることが分かる．また，初期ページについても z_6 は，最初に訪問するページが sale である確率が約 50%と非常に高い値を示している．このことから，z_6 はセールへの関心の高いユーザが集まっているといえる．

以上の購買に関係するページへの遷移や sale ページへの遷移の特徴から，潜在クラス z_6 に所属するセッションは，購買しそうな閲覧行動と類似した閲覧をするが購買に至りにくく，セール対する関心が高いため，施策実施の効果が高い値を示していたと考えられる．

5. 考察

提案モデルをクーポン発行の意思決定支援に用いる際には，リアルタイムに新規セッションに対してクーポン発行を行うべきか否かを判断する必要がある．モデルの学習によって，あらかじめパラメータから潜在クラスごとにクーポン効果が高いクラスであるか否かを定めておけば，新規セッションがどの潜在クラスに所属するのかを推定することで，クーポン発行の意思決定を行うことができる．すなわち，セッションが未終了でも，あらかじめモデルの学習とそれぞれの潜在クラスの分析ができていれば，そこまでの閲覧ページの情報を用いることで，そのセッションがどの潜在クラスへ所属するかを推定することが可能となる．

そこで本節では，新たな閲覧ページ遷移が得られた際に，そのセッションがどの潜在クラスに所属するのかを判断するための方法について述べる．いま，新規セッションに対し，最初の n 番目までの閲覧履歴 $\boldsymbol{x}_{new} = x_0^{new} x_1^{new} \cdots x_{n-1}^{new}$ が得られたとする．このとき，$\boldsymbol{x}_{new}$ が観測されたもとでのある潜在クラス $z \in \mathcal{Z}$ への事後所属確率 $P(z|\boldsymbol{x}_{new})$ を用いて，$\boldsymbol{x}_{new}$ の所属する潜在クラス $z^* \in \mathcal{Z}$ は以下のように決定される．

$$z^* = \underset{z \in \mathcal{Z}}{\operatorname{argmax}}\, P(z|\boldsymbol{x}_{new}) \tag{19}$$

また，$\boldsymbol{x}_{new}$ のある潜在クラス z への事後所属確率 $P(z|\boldsymbol{x}_{new})$ は，以下のように算出される．

$$\begin{aligned} P(z|\boldsymbol{x}_{new}) &= \frac{P(\boldsymbol{x}_{new}|z)P(z)}{\sum_{z' \in \mathcal{Z}} P(\boldsymbol{x}_{new}|z')P(z')} \\ &= \frac{P(x_0^{new}|z) \prod_{h=1}^{n-1} P(x_d^{new}|x_{h-1}^{new}, z)P(z)}{\sum_{z' \in \mathcal{Z}} P(x_0^{new}|z') \prod_{h=1}^{n-1} P(x_h^{new}|x_{h-1}^{new}, z')P(z')} \end{aligned} \tag{20}$$

実応用の際には，何ページ目までの閲覧で所属する潜在クラスを決定するのかを判断しなければならない．ここでは，式 (20) によって決められる n 番目までの閲覧履歴 $P^* = P(z^*|\boldsymbol{x}_{new})$ の潜在クラス z^* への所属確率（$P^* = P(z^*|\boldsymbol{x}_{new})$）が，閲覧ページ数 n を増やすことでどのように変化するのかを考察する．なお，学習データ中における閲覧ページ数が 10 以上のセッションを用いた．以

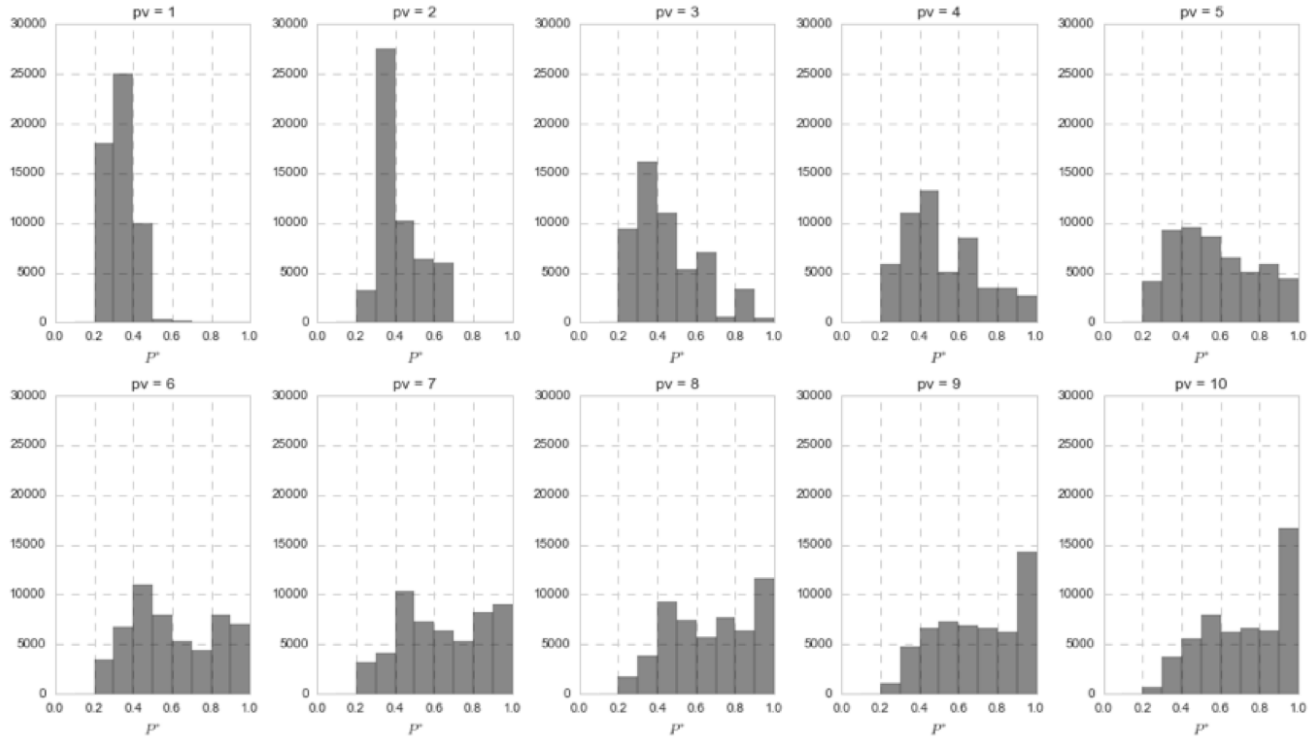


図 5 閲覧ページ数 n の変化による P^* の変化
Fig. 5 The transition of P^* by increase of page views (n).

下の**図 5** に $n = 1$〜10 まで変化させた際の，P^* のヒストグラムを示す．図 5 より，閲覧ページ数が増加するに従って，P^* の値が高いデータが増加することが分かる．閲覧ページ数が 3 となると所属確率が 0.8 以上となるセッションが現れはじめ，6 をこえると $P^* \geq 0.9$ となるセッションが増加している．このことから，閲覧ページ数 n を増やすことで，所属する潜在クラスをより正確に特定することが可能であることが明らかとなった．実際に潜在クラスを特定する際には，P^* が 0.8 をこえた時点で所属する潜在クラスを決定するなどといった閾値を設け，所属する潜在クラスの特徴に応じて施策実施すべきか否かの判断が可能である．このように，すでにモデルの学習が済んでいれば，式 (19) を用いて所属する潜在クラスをリアルタイムに決定し，その潜在クラスに応じてクーポン発行するか否かをリアルタイムに意思決定することができる．

一方，ユーザ集合全体の行動特性が変化した場合にはモデルの学習をし直す必要がある．実際の EC サイトでは，時間の経過とともに新規ユーザが追加され，退会ユーザも存在することから，ユーザが入れ替わることで，その閲覧・購買行動の統計的特徴が変化する可能性がある．そのため，定期的にモデルの整合性をチェックし，必要に応じて最新のデータが学習し直す必要がある．モデルの整合性のチェックのためには，期待コンバージョン率と実際のコンバージョン率の差異を用いることができる．実際のコンバージョン率がモデルから推定される期待値よりも低減している場合には，最新データを用いて学習し直せばよいと考えられる．

6. まとめ・今後の課題

本研究では，EC サイトにおける購買履歴に加えて，閲覧履歴，施策実施有無を考慮した確率的マルコフ潜在クラスモデルの提案を行った．また，実際の閲覧，購買履歴と割引クーポン発行施策の実施結果の実データを用い，提案モデルを適用して有効性の検証を行った．その結果から，今後の施策実施についてとるべき方向性を示すことができることが確認され，現行施策の評価を行うことで提案モデルの有用性を示した．本研究の提案モデルにより，各セッションに対して，クーポン発行の必要性を自動的に判断することができる．他の機械学習アルゴリズムで算出されたクーポン発行対象セッションに対して，本研究の提案モデルを適用し，組み合わせて用いることも容易である．すなわち，提案モデルでは何らかのクーポン発行ロジックを必要とするものの，それらと組み合わせることによってクーポン発行ロジックが効果を発揮するような潜在クラスを特定することが可能となる．したがって，クーポン発行ロジックを改良，変更したとしても新たに学習データを蓄積して学習をし直すことができれば，引き続き提案モデルと併用し続けることが可能となる．

今後の課題として，Latent Dirichelet Allocation [24] のようなパラメータをベイズ推定するようなモデルへの拡張や，異なる EC サイトおよび異なる施策についてのさらなる実験や検証などがあげられる．

謝辞 本研究にあたり，株式会社 Emotion Intelligence より，EC サイトのアクセスログデータや実験に必要な情報をご提供いただくなど，大変手厚いサポートを賜りました．本研究へのご協力に深く感謝致します．また，上智大学の山下遥氏，ならびに早稲田大学後藤研究室の方々には，本研究の内容に関して様々な議論やコメントをいただきました．さらに，本論文の査読にあたり，匿名の査読者の方々からも大変有益なコメントをいただきました．ここに深く感謝の意を表します．本研究の一部は科学研究費（26282090，26560167）の助成を受けたものである．

参考文献

[1] Radcliffe, N.J. and Surry, P.D.: Real-world Uplift Modelling with Significance-based Uplift Trees, White Paper TR-2011-1, Stochastic Solutions (2011).
[2] Hofmann, T.: Probabilistic Latent Semantic Indexing, *Proc. 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.50–57, ACM (1999).
[3] Bhatnagar, A. and Ghose, S.: A Latent Class Segmentation Analysis of e-shoppers, *Journal of Business Research*, Vol.57, No.7, pp.758–767 (2004).
[4] Bassi, F.: Latent Class Factor Models for Market Segmentation: An Application to Pharmaceuticals, *Statistical Methods and Applications*, Vol.16, No.2, pp.279–287 (2007).
[5] Ishigaki, T., Takenaka, T. and Motomura, Y.: Category Mining by Heterogeneous Data Fusion Using PdLSI Model in a Retail Service, *ICDM*, pp.857–862, IEEE Computer Society (2010).
[6] Dias, J.G. and Vermunt, J.K.: Latent Class Modeling of Website Users' Search Patterns: Implications for Online Market Segmentation, *Journal of Retailing and Consumer Services*, Vol.14, No.6, pp.359–368 (2007).
[7] Matsuzaki, Y., Mikawa, K. and Goto, M.: Modeling Customer Purchase Behavior Based on Page Transitions by Latent Class Model for Customer Segmentation, *Proc. 17th Asia Pacific Industrial Engineering and Management Systems Conference* (2016).
[8] 照井伸彦，ウィラワン・ドニ・ダハナ，伴　正隆：マーケティングの統計分析，朝倉書店 (2009).
[9] Bishop, C.: *Pattern Recognition and Machine Learning*, Springer-Verlag, New York (2006).
[10] Train, K.: *Discrete Choice Methods with Simulation*, SUNY-Oswego, Department of Economics (2003).
[11] McLachlan, G.J. and Peel, D.: *Finite Mixture Models*, Wiley Series in Probability and Statistics, J. Wiley & Sons, New York (2000).
[12] Dempster, A.P., Laird, N.M. and Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society, Series B*, Vol.39, No.1, pp.1–38 (1977).
[13] McLachlan, G. and Krishnan, T.: *The EM Algorithm and Extensions*, Vol.382, John Wiley & Sons (2007).
[14] Green, P.E., Carmone, F.J. and Wachspress, D.P.: Consumer Segmentation via Latent Class Analysis, *Journal of Consumer Research*, Vol.3, No.3, pp.170–74 (1976).
[15] Magidson, J. and Vermunt, J.K.: Latent Class Models for Clustering: A Comparison with K-means, *Canadian Journal of Marketing Research*, Vol.20, No.1, pp.36–43 (2002).
[16] Hofmann, T.: Latent Semantic Models for Collaborative filtering, *ACM Trans. Information Systems*, pp.89–115 (2004).
[17] Jin, R., Si, L. and Zhai, C.: Preference-based Graphic Models for Collaborative Filtering, *Proc. 19th Conference on Uncertainty in Artificial Intelligence, UAI'03*, pp.329–336, Kaufmann Publishers Inc. (2003).
[18] Chen, D., Wang, D., Yu, G. and Yu, F.: A PLSA-Based Approach for Building User Profile and Implementing Personalized Recommendation, *Proc. Advances in Data and Web Management, Joint 9th Asia-Pacific Web Conference, APWeb 2007, and 8th International Conference on Web-Age Information Management, WAIM 2007*, pp.606–613 (2007).
[19] Jin, R., Si, L. and Zhai, C.: A Study of Mixture Models for Collaborative Filtering, *Information Retrieval*, Vol.9, No.3, pp.357–382 (2006).
[20] Goto, M., Minetoma, K., Mikawa, K., Kobayashi, M. and Hirasawa, S.: A Statistical Model for Recommender System to Maximize Sales Amount Focusing on Characteristics of EC Site Data, *IEEE International Conference on Systems, Man, and Cybernetics*, pp.1306–1311 (2014).
[21] Yamagami, K., Fujiwara, N., Mikawa, K. and Goto, M.: A Statistical Model for Recommender System to Maximize Sales Amount Focusing on Characteristics of EC Site Data, *Proc. 15th Asia Pacific Industrial Engineering and Management Systems Conference* (2014).
[22] Goto, M., Mikawa, K., Hirasawa, S., Kobayashi, M., Suko, T. and Horii, S.: A New Latent Class Model for Analysis of Purchasing and Browsing Histories on EC Sites, *Industrial Engineering and Management Systems*, Vol.14, No.4, pp.335–346 (2015).
[23] Matsuzaki, Y., Yamagami, K., Mikawa, K. and Goto, M.: Analysis of Customer Purchase Behavior by using Purchase History with Discount Coupon Based on Latent Class Model, *Proc. 16th Asia Pacific Industrial Engineering and Management Systems Conference* (2015).
[24] Blei, D.M., Ng, A.Y. and Jordan, M.I.: Latent Dirichlet Allocation, *J. Mach. Learn. Res.*, Vol.3, pp.993–1022 (2003).

付　　録

A.1 対象データの基本統計量

以下では本研究で対象とした閲覧履歴データに対する，基本統計量を示す．なお，事前実験には，対象 EC サイトに 2016 年 4 月 1 日から 2016 年 4 月 30 日までの間に蓄積されたデータのうち，100,000 件をランダムサンプリングしたものを用いた．

購買が行われたセッション（購買セッション）と行われたかったセッション（非購買セッション）で層別を行い，それぞれ 1 セッションあたりの「平均閲覧ページ数」，「滞在時間」，「ページあたり閲覧時間」について比較行った．その結果を以下の**表 A·1**，**表 A·2** に示す．

表 A·1，A·2 より，「平均閲覧ページ数」，「滞在時間」，「ページあたり閲覧時間」のすべての項目について購買セッションの平均値，中央値が高いことが分かる．このことから，購買を行うためにある程度の時間をかけてページを閲

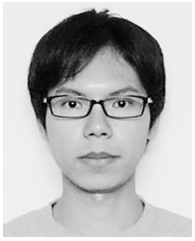
表 A·1 購買セッションの基本統計量（9,824 件）
Table A·1 Fundamental statistics of purchasing sessions.

購買セッション	閲覧ページ数	滞在時間（分）	ページあたり閲覧時間（秒）
最大値	697	1352.9	621.7
最小値	2	0	0
平均値	59.2	39.3	45.8
中央値	46	28.1	34.9
標準偏差	49.5	44.2	39.3

表 A·2 非購買セッションの基本統計量（90,174 件）
Table A·2 Fundamental statistics of non-purchasing sessions.

非購買セッション	閲覧ページ数	滞在時間（分）	ページあたり閲覧時間（秒）
最大値	850	993	16,824.9
最小値	2	0	0
平均値	16.7	11.7	48.0
中央値	8	4.3	25.9
標準偏差	24.6	24.4	130.2

覧していることが分かり，これは本研究で対象としているカタログサイトの特徴とも一致すると考えられる．また，非購買セッションの閲覧ページ数の中央値が 8 であることや滞在時間の中央値が 4 分であることから，購買に至らないセッションの半数は非常に短い時間でサイトから離脱していることが分かる．これらのことから，1 セッションにおける閲覧行動と購買の有無の間には統計的な傾向の差異が見られることが分かる．なお，本研究ではサイトを閲覧中のアクティブユーザへクーポンのリアルタイム発行を対象としていること，提案モデルとしてマルコフ性を仮定していることから，閲覧ページが複数であることが望ましい．このため，セッション数が 2 以上の顧客のデータを使用するものとした．



松嵜 祐樹

1992 年生．2017 年早稲田大学大学院修士課程修了．在学時，機械学習に基づく EC サイトにおける閲覧，購買履歴データの分析手法に関する研究に従事．

三川 健太

1981 年生．2005 年武蔵工業大学環境情報学部環境情報学科卒業．2007 年同大学大学院修士課程修了．2016 年早稲田大学大学院博士後期課程修了．博士（工学）．2013 年早稲田大学助手．2016 年湘南工科大学講師．機械学習とその応用に関する研究に従事．IEEE，電子情報通信学会，日本経営工学会等，各会員．

後藤 正幸 （正会員）

1969 年生．1994 年武蔵工業大学大学院修士課程修了．2000 年早稲田大学大学院博士課程修了．博士（工学）．1997 年早稲田大学理工学部助手．2000 年東京大学大学院工学系研究科助手．2002 年武蔵工業大学環境情報学部助教授．2008 年早稲田大学創造理工学部経営システム工学科准教授．2011 年同大教授．情報数理応用とデータサイエンスの研究に従事．著書に，『入門パターン認識と機械学習』，コロナ社 (2014)，『ビジネス統計～統計基礎とエクセル分析』，オデッセイコミュニケーションズ (2015) 等．IEEE，電子情報通信学会，人工知能学会，日本経営工学会，日本オペレーションズ・リサーチ学会，経営情報学会等，各会員．