

# EC サイトにおける施策実施効果の向上を目的とした マルコフ潜在クラスモデルに関する研究

情報数理応用研究

5215C018-5

松崎 祐樹

指導教員

後藤 正幸

## Latent Semantic Markov Model for Effective Promotion Activities in EC Sites

MATSUZAKI Yuki

### 1 研究背景・目的

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。このような施策の一例として、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挙げられる。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機会損失を招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する施策を行うことが重要である。

施策の実施が購買に繋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。例えば、Aspect Model [1] (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 [2]。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上でモデル化した Latent Segment Markov Chain (以下、LSMC) [3] が挙げられる。しかし、LSMC は閲覧履歴のみに着目したモデルであり、EC サイトにおける購買有無を考慮することができない。この LSMC に購買有無を考慮することができれば、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うに、より詳細な顧客の購買行動を記述することが可能となる。

また、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毎に異なると考えられるため、「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繋

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及び、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、LSMC 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及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 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。ユーザがサイトに訪問してから購買、サイトから離脱するまでの 1 度の閲覧行動をセッションと呼び、これに一意な ID が付与されている。この ID に閲覧履歴と購買有無が紐付けられている。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。

#### 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  $S = \{s_i : 1 \leq i \leq I\}$ 、 $J$  種類からなるページタイプ集合を  $K = \{k_j : 1 \leq j \leq J\}$  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  $\mathbf{x}_i$  は以下で与えられる。

$$\mathbf{x}_i = x_0^i x_1^i \cdots x_{T_i-1}^i \quad (1)$$

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、 $x_t^i \in K$  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 Latent Segment Markov Chain

##### 2.3.1 概要

LSMC はユーザの閲覧行動にマルコフ性を仮定し、潜在クラスモデルとして表現した統計モデルである。LSMC では、セッション  $s_i$  の持つ閲覧履歴系列  $\mathbf{x}_i$  の確率構造を潜在クラスを条件とする条件付き 1 次マルコフモデルでモデル化している。このようなモデル化を行うことによって、top ページから sale ページに遷移しやすいユーザや cart 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。

### 2.3.2 定式化

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  $\mathbf{x}_i$  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.

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

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.

## 3 閲覧履歴及び購買行動を考慮した

### マルコフ潜在クラスモデル

#### 3.1 概要

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に LSMC を拡張したモデル (以下, 購買モデル) の提案を行う. 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.

#### 3.2 定式化

提案する購買モデルでは, 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.

$$w_i = \begin{cases} 1, & s_i \text{ で購買が起る場合 (事象 } y) \\ 0, & s_i \text{ で購買が起らない場合 (事象 } \bar{y}) \end{cases} \quad (3)$$

ここで,  $s_i$  の閲覧履歴  $\mathbf{x}_i$  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  $(\mathbf{x}_i, w_i, v_i)$  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  $(\mathbf{x}_i, w_i, v_i)$  についての確率モデルは以下のように表される.

$$P(\mathbf{x}_i, w_i, v_i) = P(v_i) P(\mathbf{x}_i | v_i) P(y | v_i)^{w_i} P(\bar{y} | v_i)^{1-w_i} \quad (4)$$

なお,  $P(y | z)$  についてはある潜在クラス  $z \in \mathcal{Z}$  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  $\mathbf{x}_i$ , 購買有無  $w_i$  の生起確率  $P(\mathbf{x}_i, w_i | v_i = z_l)$  は以下ようになる.

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

ただし,  $\lambda_{lj} = P(x_0^i = k_j | z_l) = P(x_0^i = k_j | v_i = z_l)$  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.

## 3.3 施策を実施すべきクラス

購買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, 既にリアルタイムに割引クーポンが発行されており (5.1, 5.2 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. あらかじめ施策実施すべきクラスを特定しておけば, 新規セッションが得られた時点でその閲覧ページ遷移から所属する潜在クラスを特定することで, 施策実施有無の意志決定を即座に行うことが可能となる.

## 4 閲覧履歴, 購買有無及び施策実施有無を考慮した

### マルコフ潜在クラスモデル

#### 4.1 概要

前節のモデルを用いて施策実施の意志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. すなわち, 購買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言える. 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. 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, 前節の購買モデルを拡張し, 施策実施の有無を直接的に考慮したモデル (以下, 施策購買モデル) 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. 施策購買モデルのグラフィカルモデルは図 1 のように表される. なお, 図 1 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.

$$d_i = \begin{cases} 1, & s_i \text{ で施策実施された場合 (事象 } c) \\ 0, & s_i \text{ で施策実施されなかった場合 (事象 } \bar{c}) \end{cases} \quad (6)$$

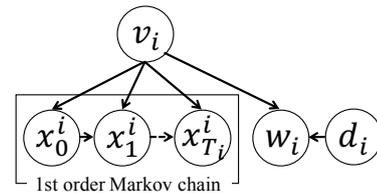


図 1: 施策購買モデルのグラフィカルモデル

## 4.2 定式化

施策購買モデルでは、購買モデルを拡張し、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。購買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  $\mathbf{x}_i$ 、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  $(\mathbf{x}_i, w_i, d_i, v_i)$  と表される。ただし、 $v_i \in \mathcal{Z} = \{z_l : 1 \leq l \leq L\}$  を満たす。よって、 $i$  番目の完全データの確率モデルは以下のように表される。

$$\begin{aligned} P(\mathbf{x}_i, w_i, v_i | d_i) \\ = P(v_i) P(\mathbf{x}_i | v_i) \{P(y|c, v_i)\}^{d_i w_i} \{P(\bar{y}|c, v_i)\}^{d_i(1-w_i)} \\ \cdot \{P(y|\bar{c}, v_i)\}^{(1-d_i)w_i} \{P(\bar{y}|\bar{c}, v_i)\}^{(1-d_i)(1-w_i)} \end{aligned} \quad (7)$$

なお、 $P(y|c, z)$  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下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  $v_i = z_l$  及び施策実施有無を表す変数  $d_i$  が与えられた下での閲覧履歴  $\mathbf{x}_i$ 、購買有無  $w_i$  の生起確率  $P(\mathbf{x}_i, w_i | v_i = z_l, d_i)$  は以下ようになる。

$$\begin{aligned} P(\mathbf{x}_i, w_i | v_i = z_l, d_i) \\ = P(\mathbf{x}_i | z_l) \{P(y|c, z_l)\}^{d_i w_i} \{P(\bar{y}|c, z_l)\}^{d_i(1-w_i)} \\ \cdot \{P(y|\bar{c}, z_l)\}^{(1-d_i)w_i} \{P(\bar{y}|\bar{c}, z_l)\}^{(1-d_i)(1-w_i)} \\ = \prod_{j=1}^J \lambda_{jl}^{\delta(x_{0j}^i = k_j)} \prod_{j=1}^J \prod_{m=1}^J a_{jml}^{n_{ijm}} \gamma_l^{d_i w_i} \bar{\gamma}_l^{d_i(1-w_i)} \\ \cdot \gamma_l^{0(1-d_i)w_i} \bar{\gamma}_l^{0(1-d_i)(1-w_i)} \end{aligned} \quad (8)$$

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

## 4.3 パラメータ推定

施策購買モデルについても潜在変数を含むため、EM アルゴリズム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】

$$\begin{aligned} P(z_l | \mathbf{x}_i, w_i, d_i) &= \frac{\pi_l P(\mathbf{x}_i, w_i | z_l, d_i)}{\sum_{l'=1}^L \pi_{l'} P(\mathbf{x}_i, w_i | z_{l'}, d_i)} \\ &= \alpha_{il} \end{aligned} \quad (9)$$

【M-step】

$$\pi_l = \frac{1}{I} \sum_{i=1}^I \alpha_{il} \quad (10)$$

$$\lambda_{jl} = \frac{\sum_{i=1}^I \alpha_{il} \delta(x_{0j}^i = k_j)}{\sum_{i=1}^I \alpha_{il}} \quad (11)$$

$$a_{jml} = \frac{\sum_{i=1}^I \alpha_{il} n_{ijm}}{\sum_{m=1}^J \sum_{i=1}^I \alpha_{il} n_{ijm}} \quad (12)$$

$$\gamma_l^1 = \frac{\sum_{i=1}^I \alpha_{il} d_i w_i}{\sum_{i=1}^I \alpha_{il} d_i} \quad (13)$$

$$\gamma_l^0 = \frac{\sum_{i=1}^I \alpha_{il} (1-d_i) w_i}{\sum_{i=1}^I \alpha_{il} (1-d_i)} \quad (14)$$

EM アルゴリズムでは、式 (15) 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。なお、 $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_I)$ 、 $\mathbf{W} = (w_1, \dots, w_I)$ 、 $\mathbf{D} = (d_1, \dots, d_I)$ 、 $\mathbf{V} = (v_1, \dots, v_I)$  である。

$$\log P(\mathbf{X}, \mathbf{W}, \mathbf{D}, \mathbf{V}) = \sum_{i=1}^I \log P(\mathbf{x}_i, w_i, d_i, v_i) \quad (15)$$

## 5 実験

### 5.1 概要

施策購買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、閲覧履歴及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、既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。

### 5.2 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、A/B テストが行われている。この A/B テストでは、サイトを訪れた全てのセッションに対してランダムに 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る。そこで本実験では、クーポン対象の有無と A/B 群の観点からデータに表 1 のような 4 つのラベル付けを行う。

表 1: データのラベル付け

		クーポン対象有無	
		有	無
購買有無	有	A1	B1
	無	A0	B0

このラベルを用い、モデル学習により推定された A1 群の購買確率と B1 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施策購買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 A1, B1 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。

### 5.3 実験条件

実験には、2016 年 4 月 1 日から 30 日までの 1ヶ月間で蓄積された閲覧、購買履歴及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 A1, B1）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う、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、施策購買モデルでは、クーポン対象セッション（ラベル A1, B1）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、閲覧端末がPCであるデータのみを用い、潜在変数の数は事前実験より  $L = 10$  とした。

#### 5.4 実験結果・考察

A1群の購買確率  $P(y|z, c)$  と B1群の購買確率  $P(y|z, \bar{c})$  の値や施策実施効果、カバー率に着目し、分析を行った。表2に特徴的な結果を示した3つの潜在クラス ( $z_3, z_5, z_6$ ) の結果を示す。なお、施策実施効果は、A1群とB1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 A1, B1) となったセッションの割合によって算出される。

表2: 実験結果

潜在クラス	$z_3$	$z_5$	$z_6$	全体
所属セッション数	8,402	18,951	17,226	109,708
混合比	0.08	0.17	0.16	1.00
A1の購買確率	0.088	0.045	0.263	0.142
B1の購買確率	0.060	0.040	0.225	0.117
施策実施効果 (A1/B1)	1.46	1.12	1.17	1.21
カバー率	45%	58%	59%	50%

表2より、 $z_6$  について A1, B1 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。そのため、このクラスへのクーポン発行は余分な割引を行うことになるため、施策を実施すべきではない。一方  $z_3$  は、B1の購買確率が低い値を取っているが、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言えるため、重点的に施策を実施すべきである。また、 $z_5$  については、B1の購買確率、施策実施効果ともに低い値を取っている。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言える。

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%と高い値を取っている。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言える。また、購買意欲が低いと考えられる  $z_5$  のカバー率は58%と高い値を取っていることから、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。

#### 6 クーポン発行タイミングに関する考察

クーポン購買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。

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  $\mathbf{x}_{new} = (x_0^{new}, x_1^{new}, \dots, x_{n-1}^{new})$  が得られたとする。このとき、 $\mathbf{x}_{new}$  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  $P(z|\mathbf{x}_{new})$  を用いて、 $\mathbf{x}_{new}$  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。

$$z^* = \arg \max_{z \in \mathcal{Z}} P(z|\mathbf{x}_{new}) \quad (16)$$

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

$$\begin{aligned} P(z|\mathbf{x}_{new}) &= \frac{P(\mathbf{x}_{new}|z)P(z)}{\sum_{z' \in \mathcal{Z}} P(\mathbf{x}_{new}|z')P(z')} \\ &= \frac{P(x_0^{new}|z) \prod_{h=1}^{n-1} P(x_h^{new}|x_{h-1}^{new}, z)P(z)}{\sum_{z' \in \mathcal{Z}} P(x_0^{new}|z') \prod_{h=1}^{n-1} P(x_h^{new}|x_{h-1}^{new}, z')P(z')} \end{aligned} \quad (17)$$

このように、既にモデル学習が済んでいれば、式(16)を用いて所属する潜在クラスを瞬時に決定し、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。

#### 7 まとめ・今後の課題

本研究では、ECサイトにおける閲覧履歴に加えて、購買有無を考慮したモデル及び、購買有無と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を用い、クーポン購買モデルによる実験を行った。その結果から、今後の施策実施について取るべき方向性を示し、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。今後の課題として、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が挙げられる。

#### 参考文献

- [1] T. Hoffman, "Probabilistic Latent Semantic Indexing," *Proc. the 22nd Annual Int. SIGIR Conf. Research and Development in Information Retrieval*, pp. 50–57, 1999.
- [2] A. Bhatnagar and S. Ghose, "A Latent Class Segmentation Analysis of E-shoppers," *J. Business Research*, Vol. 57, No. 7, pp. 758–767, 2004.
- [3] J. G. Dias and J. K. Vermunt, "Latent Class Modeling of Website Users' Search Patterns: Implications for Online Market Segmentation," *J. Retailing and Consumer Services*, No. 14, pp. 359–368, 2007.
- [4] A. Dempster, N. Laird and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. the Royal Statistical Society, Series B*, Vol. 39, No. 1, pp. 1–38, 1977.