

会員ステージ間の潜在クラス分布の差異を検出する購買行動 分析手法に関する一考察

楊 添翔^{1,a)} 山下 遥² 後藤 正幸¹

概要: 年間累積購買金額がある額を超えると顧客のステージが上位へ向上し、ポイントなどの特典を提供する会員ステージ制というマーケティング施策がある。この仕組みにより、企業側は、顧客の購買意欲を高めるだけでなく、その購買履歴データを得て優良顧客特定などの分析に用いることが可能である。例えば、会員ステージ間での購買特性の差異を明らかにするため、会員ステージ毎にクラスタリングモデルを構築することは顧客理解のために有用である。しかし、ステージ間で独立にクラスタリングモデルを学習した場合、得られたクラスを会員ステージ間で比較分析することができない。本研究では、会員ステージ間でクラス分布の差異が分析できるモデルとその学習アルゴリズムを提案する。

TIANXIANG YANG^{1,a)} HARUKA YAMASHITA² MASAYUKI GOTO¹

1. はじめに

近年、情報技術の発展に伴い、多くの企業は顧客の行動に関する様々なデータを大量に蓄積できるようになっている。これにより、多くの企業が蓄積した ID-POS データをどのようにビジネス上の施策に活用するのかについて議論が重ねられている [1][2]。しかし、企業によって分析ニーズも異なり、万能の分析手法が存在する訳ではないため、新たな特性を持つデータの生成によってデータの活用方法を変える必要がある。本研究では、小売業において顧客の優良性を判別するため、会員のある期間の累積購買金額に対応して、顧客のステージランクを決める会員ステージシステムを対象とする。このシステムの下で、期間内の累積購買金額がある額を超えると顧客のステージランクが上位へ向上し、ポイントやクーポンなどの特典を提供する。このシステムの運用により、企業が、優良顧客を特定すると共に、その購買履歴データを得て分析に用いることが可能となる。本研究では、会員ステージランク制度を導入した企業を分析対象とする。このシステムを導入する企業にとって、会員ステージ毎の顧客の購買行動を把握することは重要なため、マーケティング分野の分析ではこのような顧客

の異質性を考慮したモデルが用いられることが多い。潜在クラスモデルの 1 つである Probabilistic Latent Semantic Analysis[3] (以下、PLSA) モデルはその代表的な手法としてよく使われる。しかし、従来手法である PLSA を会員ステージ制を採用している小売の購買履歴データに適用する場合、会員ステージ毎に独立に PLSA の学習を行うと、それぞれ全く異なる潜在クラスを構成してしまうため、会員ステージ間の購買行動差異を分析することが難しいという問題が生じる。そこで、本研究では、全ステージの顧客全体から潜在クラスを推定し、購買アイテムに対して同じ特性を持つ潜在クラス集合上で各ステージの顧客の所属分布を推定することで、各ステージにおける潜在クラス分布の差異を比較可能な PLSA モデルを提案する。提案手法により、人工データと実データを用いた分析の結果を示し、提案手法の有用性を示す。

2. 準備

2.1 ステージランク制度システム

ステージランク制度システムとは会員顧客のある期間の累積購買金額に対応して、会員ステージを定義し、新たな会員ステージ到達する毎に、会員にクーポンやショッピングポイントなどの特典を付与するシステムである。このシステムでは、一般的に会員顧客の累積金額が高ければ高いほど、付与される特典の価値が高くなるため、顧客に対して購買を促すインセンティブを与えることが期待できる。

¹ 早稲田大学
Waseda University, Shinjyuku, Tokyo 169-0072, Japan

² 上智大学
Sophia University

a) you.tensyou@akane.waseda.jp

ただし、企業によって顧客の優良性について解釈が異なるため、設定される特典とステージランクの対応方法も異なる。本研究の対象となる企業では付与された特典は一定期間を過ぎると消滅し、顧客のステージランクをリセットし、ステージを再スタートさせるという特徴もある。近年、ビジネスに対する顧客分析の重要性を高めるため、ID-POSだけでなく、様々な企業はこのステージランク制度を会社の購買システムに導入している企業は少なくない。例えば、Rakuten、ファミリーマートやドン・キホーテなどの会社がこのシステムを利用している。このシステムは顧客の購買意欲を高めるだけではなく、購買の履歴データが会員ステージの情報と共に蓄積されるため、顧客の購買行動や優良性についての分析も可能となる。

一方で、ステージランク制度システムの下で顧客の購買行動データには、以下の二つの特性がある：

- (1) 顧客のステージランクが高くなると、商品の購買範囲が広くなり、購買パターンも複雑になる。
- (2) ステージランクが高ければ高いほど、顧客数が少なくなる。

企業側がビジネス利益を求めため、ステージランクが高い顧客層の購買行動の分析を重視している。従って、有効な手法で人数が少ない高ステージのデータからビジネスのポイントを発見するのは重要な課題である。これまでもいくつかの研究が行われているが [4][5][6][7][8]、ステージ間で顧客クラスを比較することを目的とした研究は行われていない。

2.2 Probabilistic Latent Semantic Analysis[3]

PLSA は潜在クラスモデルの一つである。このモデルはユーザとアイテムの間いくつかの潜在クラスを仮定し、潜在クラスの下でユーザがアイテムを購入する事象をユーザとアイテム共起で表現する確率潜在クラスモデルである。EC サイトにおける推薦システムや顧客のクラスターリングなど様々な分野にも応用されている [9][10][11]。PLSA のグラフィカルモデルを図 1 に示す。

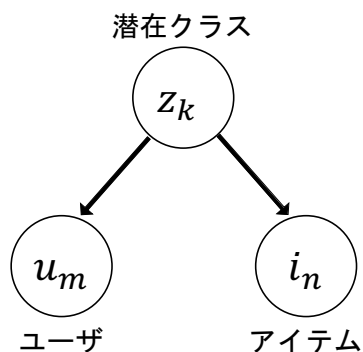


図 1 PLSA のグラフィカルモデル
 Fig. 1 Graphical model of PLSA

ここで、潜在クラスの下でユーザがアイテムを購入する事象を共起で表現するため、 M 人のユーザ集合を $U = \{u_m : 1 \leq m \leq M\}$ 、 N 個のアイテム集合 $I = \{i_n : 1 \leq n \leq N\}$ 、 K 個の潜在クラス集合を $Z = \{z_k : 1 \leq k \leq K\}$ と定義する。PLSA の確率モデルは式 (1) で表される。

$$P(u_m, i_n) = \sum_{k=1}^K P(z_k)P(u_m|z_k)P(i_n|z_k), \quad (1)$$

ここで、観測不可能な潜在変数 z_k が存在するため、式 (1) のパラメータ $P(z_k)$ 、 $P(u_m|z_k)$ 、 $P(i_n|z_k)$ は EM アルゴリズム [12][13] により推定する。本研究の対象事例では、ステージ毎にこのモデルを作成することが可能である。しかしながら、ステージ毎のモデルは全て独立に推定されるため、得られたパラメータの比較が困難となるという課題が存在する。

3. 提案手法

3.1 概要

前節で述べたように、PLSA などの潜在クラスモデルを実データを適用する際、ステージ毎に独立にモデルを学習した場合、得られた潜在クラスを会員ステージ間で比較分析することが困難である。これに対して、全ステージの顧客の購買履歴データを 1 つの潜在クラスモデルでまとめて学習するというアプローチも考えられるが、その場合、得られた潜在クラスには各ステージのユーザが混在するため、ステージ間の顧客購買行動の変化を考察することが困難である。上記の問題点を解決するため、本研究では、ステージ間のユーザの購買行動の差異に着目し、ステージ間で潜在クラスの所属割合がどのように異なるかを比較可能なモデルとその学習法を提案する。提案モデルでは、まず、真の潜在クラス数は観測できないため、全ステージの数と各ステージの潜在クラスの数と同じであることを仮定する。加えて、ステージ間における潜在クラス毎のユーザの出現分布を比較するため、全ステージの潜在クラス下でのアイテムの出現確率と各ステージの潜在クラス下でのアイテムの出現確率が同一であることを仮定する。これにより、各ステージで潜在クラスの購買特徴が同一になり、同じ潜在クラス集合上で、ステージ間での出現確率の違いを比較分析可能である。

まず、全てのステージデータを用い、全ユーザの購買傾向モデルを構築する。そこから推定された全体の潜在クラス下でのアイテムの出現確率を各ステージのモデルの初期値として導入する。そして、各ステージの EM アルゴリズムの推定段階で、導入された初期値を固定しながら、モデルのパラメータを推定する。

ここで、潜在クラスの下でユーザがアイテムを購入する

事象を共起の事象として表現するため、全ステージの総合 M 人のユーザ集合を $\mathcal{U} = \{u_m : 1 \leq m \leq M\}$, N 個のアイテム集合を $\mathcal{I} = \{i_n : 1 \leq n \leq N\}$, K 個の潜在クラス集合を $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$ と定義する。全ステージの中に、 s 段階のステージを設定し、 $s \in \mathbb{N}$ と表す。ステージ s に M^s 人のユーザ集合を $\mathcal{U}^s = \{u_{m^s} : 1 \leq m^s \leq M^s\}$ と設定する。ここで、 $M = M^1 + M^2 + \dots + M^s$, $\forall \mathcal{U}^s \subseteq \mathcal{U}$ である。

3.2 全ステージデータにおけるユーザとアイテム共起モデル

全ステージユーザモデルでは、ユーザ u_m がアイテム i_n を購入した事象を (u_m, i_n) と記述する。この時、モデルは以下の式 (2) のようにで表すことができる。

$$P(u_m, i_n) = \sum_{k=1}^K P(z_k)P(i_n|z_k)P(u_m|z_k), \quad (2)$$

学習データに対する対数尤度関数 LL は以下の式 (3) のように示される。

$$LL = \sum_{m=1}^M \sum_{n=1}^N \left\{ \delta(u_m, i_n) \times \sum_{k=1}^K P(z_k|u_m, i_n) \log P(z_k)P(i_n|z_k)P(u_m|z_k) \right\}. \quad (3)$$

ただし、 $\delta(u_m, i_n)$ は、ユーザ u_m がアイテム i_n を購入していた場合 1, そうでない場合 0 を返す指示関数である。この対数尤度関数を最大化するパラメータは、EM アルゴリズムにより、推定する。

パラメータ更新ステップは以下のように示される：

Step 1 : 各パラメータ $P(z_k)$, $P(i_n|z_k)$, $P(u_m|z_k)$ の初期値は乱数を用いて生成する。

Step 2 : EM アルゴリズムにより、パラメータを更新する。

[E-Step]:

$$P(z_k|u_m, i_n) = \frac{P(z_k)P(i_n|z_k)P(u_m|z_k)}{\sum_{k=1}^K P(z_k)P(i_n|z_k)P(u_m|z_k)}, \quad (4)$$

[M-Step]:

$$P(u_m|z_k) = \frac{\sum_{n=1}^N P(z_k|u_m, i_n)\delta(u_m, i_n)}{\sum_{m=1}^M \sum_{n=1}^N P(z_k|u_m, i_n)\delta(u_m, i_n)}, \quad (5)$$

$$P(i_n|z_k) = \frac{\sum_{m=1}^M P(z_k|u_m, i_n)\delta(u_m, i_n)}{\sum_{m=1}^M \sum_{n=1}^N P(z_k|u_m, i_n)\delta(u_m, i_n)}, \quad (6)$$

$$P(z_k) = \frac{\sum_{m=1}^M \sum_{n=1}^N P(z_k|u_m, i_n)\delta(u_m, i_n)}{\sum_{k=1}^K \sum_{m=1}^M \sum_{n=1}^N P(z_k|u_m, i_n)\delta(u_m, i_n)}. \quad (7)$$

Step 3 : Step 2 により更新された対数尤度関数 LL を一定の条件で判定する。

3.3 各ステージにおけるユーザ比較モデル構築

各ステージのユーザモデルでは、ユーザ u_{m^s} がアイテム i_n を購入した事象を (u_{m^s}, i_n) と記述する。本研究で対象とする事例では、「ユーザは必ず 1 つのステージに属する」ものとする。よって、各ステージのユーザは重複することがない。一方で、各ステージのユーザが違うが、アイテムは全てのステージで共通である。そこで、各ステージでのユーザ出現確率が比較可能なモデルを作るため、全ステージの潜在クラス下でのアイテムの出現確率と各ステージの潜在クラス下でのアイテムの出現確率が一定になるものと仮定する。従って、3.2 節で推定された全ステージモデルから推定されたパラメータ $P(i_n|z_k)$ を各ステージモデルの初期値として各ステージのモデルに導入する。この時、モデルは以下の式 (8) のようにで表現することができる。

$$P(u_{m^s}, i_n) = \sum_{k=1}^K P_s(z_k)P(i_n|z_k)P(u_{m^s}|z_k), \quad (8)$$

また、学習データに対する対数尤度関数 LL^s は以下の式 (9) のように示される。

$$LL^s = \sum_{m=1}^M \sum_{n=1}^N \left\{ \delta(u_{m^s}, i_n) \times \sum_{k=1}^K P(z_k|u_{m^s}, i_n) \log P_s(z_k)P(i_n|z_k)P(u_{m^s}|z_k) \right\}. \quad (9)$$

ただし、 $\delta(u_{m^s}, i_n)$ は、ユーザ u_{m^s} がアイテム i_n を購入していた場合 1, そうでない場合 0 を返す指示関数である。ここで、全ステージの顧客のデータを用いて推定したい潜在クラス毎のアイテムの生起確率を固定しながら、各ステージの顧客の潜在クラスにおける出現確率を推定するため、以下のような EM アルゴリズムを提案する。

各ステージのパラメータ更新ステップは以下のようになる：

Step 1:各パラメータ $P_s(z_k)$, $P(u_{m^s}|z_k)$ の初期値は乱数を用いて生成する。

Step 2:EM アルゴリズムにより、パラメータを更新する。

[E-Step]:

$$P(z_k|u_{m^s}, i_n) = \frac{P_s(z_k)P(i_n|z_k)P(u_{m^s}|z_k)}{\sum_{k=1}^K P_s(z_k)P(i_n|z_k)P(u_{m^s}|z_k)}, \quad (10)$$

[M-Step]:

$$P(u_m^s | z_k) = \frac{\sum_{n=1}^N P(z_k | u_m^s, i_n) \delta(u_m^s, i_n)}{\sum_{m^s=1}^{M^s} \sum_{n=1}^N P(z_k | u_m^s, i_n) \delta(u_m^s, i_n)}, \quad (11)$$

$$P(z_k) = \frac{\sum_{m^s=1}^{M^s} \sum_{n=1}^N P(z_k | u_m^s) \delta(u_m^s, i_n)}{\sum_{k=1}^K \sum_{m^s=1}^{M^s} \sum_{n=1}^N P(z_k | u_m^s, i_n) \delta(u_m^s, i_n)}. \quad (12)$$

Step 3 : 収束判定条件は 3.2 節と同様である。

4. 実験 1

本研究では、提案手法の有効性を示すために、人工データを生成し、シミュレーションを行った。

4.1 人工購買履歴データの生成手順 [14]

ここで、ユーザ、アイテムをそれぞれ G 個のグループに分け、同じグループに所属するユーザは所属グループのアイテムを購入し易い傾向がある（似た購買傾向を持つ）と仮定する。そのために、予めユーザとアイテムを一定の数に G 分割し、 G 個のユーザ・アイテムグループを作成しておく。このもとで、以下のように、購買履歴の人工データを生成する。

具体的な手順は以下のようになる：

[Step 1]: グループの数 (G 個) と各グループの出現確率を設定する。

[Step 2]: 設定されたグループから α ($0 \leq \alpha \leq 1$) の確率でユーザ u_m を、それ以外のグループを $\frac{1-\alpha}{G-1}$ の確率で生成する。

上記の Step2 で選択されたグループに所属するアイテムの中から β ($0 \leq \beta \leq 1$) の確率でランダムにアイテムを 1 つ選択し、購買アイテムとする。それ以外のグループのアイテムを $\frac{1-\beta}{G-1}$ の確率で選択する。すなわち、ユーザ u_m がそのアイテムを購入したという結果をデータに加え、生成データの最大数に達していなければ Step 1 に戻る。

以上の Step により、各ユーザは、自身が所属する一つのグループのアイテムに対して購買傾向が強いが、それ以外のグループのアイテムについてもある程度の割合で購入するような購買履歴データが生成される。ただし、パラメータ α はユーザ自身があるグループに所属される確率を表すパラメータであり、これが大きいと“あるカテゴリに所属するユーザが、このカテゴリ中のアイテムのみを購入する傾向”が強い購買履歴データとなり、逆にこれが小さいと“あるカテゴリに所属しないユーザが他のアイテムのみも購入する傾向”がある購買履歴データとなる。これらの各グループが潜在クラスモデルの各潜在クラスに対応する。

る。人工購買履歴データ生成のイメージを図 2 に示す。これらの各グループが潜在クラスモデルの各潜在クラスに対応する。

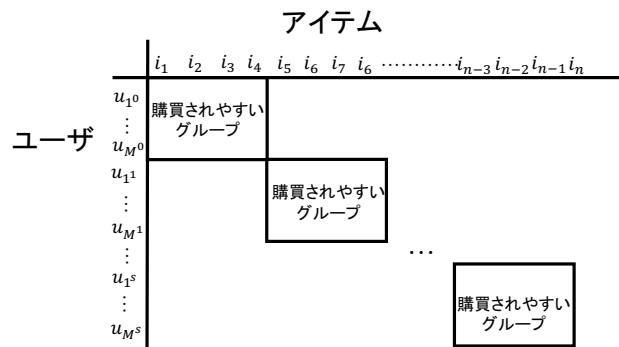


図 2 人工購買履歴データ生成のイメージ

Fig. 2 The visualization of artificial purchase history data's generation produce

4.2 シミュレーション

本稿では、ステージの成長により、グループ（潜在クラス）の出現確率が変化する状況を真のモデルとして設定する。具体的には、表 1 ように各潜在クラスの特徴を設定し、人工購買履歴データを生成する。ここで、ステージ数は 0 から 4 まで 5 段階、グループ数 $G = 5$ である。潜在クラス z_2, z_3, z_4 は購買変化が小さいグループ、 z_1 がステージの成長により出現確率が小さくなるグループ、 z_5 がステージの成長により出現確率が大きくなるグループと設定する。また、 $\alpha=0.9, \beta=0.9$, アイテム種類数 1000. ステージごとのユーザ数とアイテム平均購買数は表 2 のように設定する。生成した人工データに対し、提案手法を実行する。実験結果が比較しやすいため、各ステージの $P(z_2)$ の初期値を $1/G$ で設定する。

表 1 真の $P(z)$ の設定

Table 1 The setting of real $P(z)$

	$P(z_1)$	$P(z_2)$	$P(z_3)$	$P(z_4)$	$P(z_5)$
Stage 4	0.6	0.1	0.1	0.1	0.1
Stage 3	0.5	0.1	0.1	0.1	0.2
Stage 2	0.4	0.15	0.1	0.05	0.3
Stage 1	0.3	0.1	0.1	0.1	0.4
Stage 0	0.2	0.05	0.1	0.15	0.5

4.3 結果

提案手法で人工購買履歴データの結果を表 3 に示す。ただし、 \hat{P} は推定された確率（パラメータ）であることを示す。表 1-3 を比較すると、毎回推定される潜在クラスが異なるが、パラメータ $\hat{P}(i_n | z_k)$ が固定されたことで、提案手

表 2 ステージごとのユーザ数とアイテム平均購買数の設定

Table 2 Setting of users and item average purchase number for each stage

ステージ名	人数 (人)	アイテム平均購買数
Stage 4	1,000	160
Stage 3	3,000	80
Stage 2	9,000	40
Stage 1	27,000	20
Stage 0	81,000	10

法で推定された潜在クラスの出現確率と設定された潜在クラス (グループ) の真の確率の傾向が似ている。従って、提案手法の有効性を証明した。

表 3 人工データ購買履歴に対する推定結果

Table 3 The result of artificial purchase history data

	$\hat{P}(z_1)$	$\hat{P}(z_2)$	$\hat{P}(z_3)$	$\hat{P}(z_4)$	$\hat{P}(z_5)$
Stage 4	0.5665	0.0082	0.1459	0.0376	0.2419
Stage 3	0.4791	0.0155	0.1293	0.0397	0.3363
Stage 2	0.4314	0.0214	0.1223	0.0239	0.4010
Stage 1	0.3470	0.0281	0.1065	0.0399	0.4785
Stage 0	0.2595	0.0342	0.0902	0.0553	0.5609

5. 実験 2

5.1 分析対象データ

本稿では、株式会社良品計画から提供頂いた会員登録購買履歴データを分析対象とする。無印良品のステージと顧客年間累計購買金額概要は表 4 のようになっている。分析の対象期間は 2014 年 3 月 1 日～2015 年 2 月 28 日である。本稿では、商品アイテムが 689 種類の商品カテゴリ (仮に、「食品 A」、「食品 B」などと示す) に分類されており、本研究ではこの商品カテゴリを用いて分析を行う。

表 4 無印のステージと顧客年間累計購買金額概要

Table 4 Relationship of Muji's stage between customer's yearly purchase amount

	年間累積金額	特典	人数 (人)
Stage 4(ダイヤモンド)	20 万円以上	0	9,470
Stage 3(プラチナ)	10 万円以上	200points	20,977
Stage 2(ゴールド)	5 万円以上	300points	46,726
Stage 1(シルバー)	2 万円以上	500points	125,939
Stage 0(ステージなし)	2 万円未満	1000points	593,321

5.2 提案モデルの推定結果

事前分析により、全ユーザのモデルと各ステージの潜在クラス数は $K = 5$ とした。提案手法で推定されたパラメータ $\hat{P}(z_k)$ の推定結果を表 5 を用いて示す。ステージから得られるパラメータ $\hat{P}(i_n|z_k)$ から、ベイズ定理により、 $\hat{P}(z_k|i_n)$ を計算する。 $\hat{P}(z_k|i_n)$ の推定結果を表 6 で示す。

ここで、 $\hat{P}(z_k)$ は各ステージの潜在クラス出現確率の推定値である。 $\hat{P}(z_k|i_n)$ はあるアイテムの下で各クラスの出現確率である。

表 5, 6 より、各ステージのユーザの嗜好の特徴を把握することができる。

- (1) 全体から見ると、ステージが成長すれば、顧客の購買範囲が広がってゆくことが分かった。
- (2) クラス 2 の代表的な商品は衣服・雑貨と食品である。
- (3) クラス 3 の代表的な商品は衣服・雑貨と生活雑貨である。
- (4) ステージランクが高くなると、衣服・雑貨と食品を購入する事象の割合が大幅に減少することが分かる。
- (5) 衣服・雑貨と生活雑貨を購入するという事象の割合も少なくなる。
- (6) クラス 1, 4 では、ステージ間で出現確率はほぼ一定であり、購買嗜好の変化が小さい。
- (7) 多様なカテゴリのアイテムを購入するユーザが、各ステージに一定の割合で存在することが分かった。
- (8) クラス 5 では、ステージが高ければ高いほど、所属する人数共起事象の割合が増加することで、このクラスはステージが高い会員顧客が好みの代表商品からなるクラスと解釈することができる。すなわち、このクラスに所属する商品を購入してもらえようような施策を施すことによって会員の顧客の成長を促すような商品であるということが示唆される。

表 5 全ステージデータにおけるユーザとアイテム共起モデルの潜在クラス

Table 5 Latent class of user and item co-occurrence model by all stage data

	$\hat{P}(z_1)$	$\hat{P}(z_2)$	$\hat{P}(z_3)$	$\hat{P}(z_4)$	$\hat{P}(z_5)$
Stage 4	0.1229	0.1156	0.2475	0.1189	0.3952
Stage 3	0.1399	0.1203	0.2691	0.1073	0.3634
Stage 2	0.1219	0.1352	0.3484	0.1181	0.2764
Stage 1	0.1134	0.2026	0.2877	0.1643	0.2320
Stage 0	0.1837	0.2685	0.3736	0.1742	0.0001

6. 考察

提案手法では全ステージの顧客データから推定したモデルのパラメータ $\hat{P}(i_n|z_k)$ を各ステージのモデルに導入し、各ステージのモデルを構築した。従来の PLSA に代わり、各ステージのモデルを学習する際、EM アルゴリズムの段階で、パラメータ $\hat{P}(i_n|z_k)$ が固定されたことで、推定された各ステージのモデルパラメータは全ステージのモデルに合わせるため、各ステージの購買行動の変動は比較できるようになった。

提案手法では、全ステージモデルと各ステージモデルにおけるユーザとアイテム共起モデルを設定しており、パラ

表 6 各ステージの代表的なアイテムの出現確率ランキング
Table 6 Appearance probability ranking of representative items for each stage

順位	アイテム名	$\hat{P}(z_1 i_n)$	アイテム	$\hat{P}(z_2 i_n)$	アイテム	$\hat{P}(z_3 i_n)$	アイテム	$\hat{P}(z_4 i_n)$	アイテム	$\hat{P}(z_5 i_n)$
1 位	衣服雑貨 1	1.0000	衣服雑貨 6	0.9999	生活雑貨 5	1.0000	衣服雑貨 14	1.0000	生活雑貨 11	1.0000
2 位	生活雑貨 1	0.9998	衣服雑貨 7	0.9997	衣服雑貨 9	0.9996	衣服雑貨 15	0.6136	多角化商品 2	1.0000
3 位	生活雑貨 2	0.6362	衣服雑貨 8	0.6471	衣服雑貨 10	0.9940	多角化商品 1	0.5712	衣服雑貨 21	1.0000
4 位	生活雑貨 3	0.5982	食品 2	0.6412	生活雑貨 6	0.8842	衣服雑貨 16	0.5587	生活雑貨 12	1.0000
5 位	衣服雑貨 2	0.5717	食品 3	0.5552	衣服雑貨 11	0.8344	サービス 1	0.5364	衣服雑貨 22	1.0000
6 位	食品 1	0.5510	衣服雑貨 4	0.4920	生活雑貨 7	0.8330	生活雑貨 10	0.5314	生活雑貨 13	1.0000
7 位	生活雑貨 4	0.5339	生活雑貨 4	0.4494	衣服雑貨 12	0.8242	衣服雑貨 17	0.5165	多角化商品 3	1.0000
8 位	衣服雑貨 3	0.5247	食品 4	0.3955	衣服雑貨 13	0.8049	衣服雑貨 18	0.5162	生活雑貨 14	1.0000
9 位	衣服雑貨 4	0.5080	食品 5	0.3659	生活雑貨 8	0.7766	衣服雑貨 19	0.4768	衣服雑貨 23	0.9985
10 位	衣服雑貨 5	0.4986	食品 6	0.3634	生活雑貨 9	0.7750	衣服雑貨 20	0.4764	衣服雑貨 24	0.8996

メータ $\hat{P}(i_n|z_k)$ の推定が可能である。それとベイズ定理によって、 $\hat{P}(z_k|i_n)$ の推定もできる。それで、各クラスの代表的な商品の発見ができる。さらに、各ステージの購買行動の変動によって、ステージに成長に役を立つ商品を発見することもできる。そのため、企業側からの観点から、顧客のステージランクの向上を促す施策を考えることが可能である。

7. まとめと今後の課題

本研究では、会員ステージ間でクラス分布の差異を比較分析できるモデルとその学習アルゴリズムを提案した。提案モデルでは、全ステージから推定されたパラメータを再利用し、各ステージのモデルを構築する。人工データでのシミュレーションと実データの分析結果より、顧客の購買行動と到達ステージとの関係性を明らかにすることが可能であることを示した。さらに、得られる結果を用い、ビジネス上の施策の可能性について検討した。

本研究のモデルでは、会員ステージ間で、顧客セグメントのサイズがどのように変化しているかを比較分析することが可能であり、様々な施策の検討へと結び付けることができる。今後の課題として、ステージランク高いユーザの購買行動の要素を再利用し、低いランクステージの顧客を成長されるモデルを検討することなどが挙げられる。また、異なるステージランクシステムのデータについて実験や検証などがあげられる。

謝辞 本研究に際し、ご協力を頂いている株式会社良品計画に深く感謝いたします。ここに深く感謝の意を表します。

参考文献

[1] 岩田具治, 渡部晋治, 山田武士, 上田修功: 購買行動解析のためのトピック追跡モデル, 電子情報通信学会論文誌 D, Vol. 93, No. 6, pp. 978-987 (2010).
[2] 石垣司, 竹中毅, 櫻井瑛一, 本村陽一: 行動履歴データとライフスタイル調査にもとづく顧客モデル構築技術, 電気学会論文誌 C (電子・情報・システム部門誌), Vol. 133,

No. 9, pp. 1787-1795 (2013).
[3] Hofmann, T.: Probabilistic latent semantic analysis, *In Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc., pp. 289-296 (1999).
[4] 酒井拓哉, 三川健太, 後藤正幸: 会員ステージに着目した優良顧客・離反顧客の特徴分析, 経営システム, Vol. 25, No. 3, pp. 182-187 (2015).
[5] Nishio, Y., Itou, H., Yamashita, H. and Goto, M.: A New Analytical Model for Customer Growth Considering Potential Purchasing Preferences, *15th Asian Network for Quality Conference (ANQ2017)*, Soaltee Crowne Plaza, Kathmandu, Nepal, NKT-05 (2017).
[6] Itou, H., Kumoi, G. and Goto, M.: A Study on Extraction of Important Items Focused on Customer Growth Based on Network Analysis, *The 18th Asia Pacific Industrial Engineering and Management System Conference (APIEMS2017)*, Yogyakarta, Indonesia, ID-164 (2017).
[7] Yang, T., Kumoi, G., Yamashita, H., Goto, M.: Transfer Learning Based on Probabilistic Latent Semantic Analysis for Analyzing Purchase Behavior Considering Customers' Membership Stages, *2017 Asian Conference of Management Science & Applications (ACMSA2017)*, China, Fujian, ID-164 (2017).
[8] 楊添翔, 雲居玄道, 山下遙, 後藤正幸: 会員ステージ制における顧客の行動分析のための転移学習モデルについて, 日本経営工学会春季大会予稿集, K11 (2018 年)
[9] 清水良太郎, 坂元哲平, 山下遥, 後藤正幸: アンケートデータを考慮した EC サイトの購買履歴分析モデルの提案, 経営システム, Vol. 27, No. 2, pp. 70-76 (2017).
[10] 石垣司, 竹中毅, 本村陽一: 日常購買行動に関する大規模データの融合による顧客行動予測システム, 人工知能学会論文誌, Vol. 26, No. 6, pp. 670-681 (2011).
[11] 松寄祐樹, 三川健太, 後藤正幸: 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, 情報処理学会論文誌, Vol. 58, No. 12, pp. 2034-2045 (2017).
[12] Dempster, A. P., Laird, N.M. and Rubin, D.B.: Maximum likelihood from incomplete via the EM algorithm, *Journal of the Royal Statistical Society. Series B(Methodological)*, Vol. 39, No. 1, pp. 1-38 (1977).
[13] 宮川雅巳: EM アルゴリズムとその周辺, 応用統計学, Vol. 16, No. 1, pp. 1-21 (1987).
[14] 藤原直広, 三川健太, 後藤正幸: 閲覧及び購買行動を同時に表現するアスペクトモデルによる購買予測手法の提案, 経営情報学会誌, Vol. 26, No. 1, pp. 1-16 (2017).