

# 潜在的な購買傾向を考慮した顧客の会員ステージ向上策に関する研究

1X13C087-2 西尾 友里  
指導教員 後藤 正幸

## 1 研究背景と目的

近年、多くの小売業では、購買履歴データを解析することで、顧客の購買傾向や嗜好を把握し、売上金額増加のための様々な施策を講じている。小売店舗の売上金額を増加させるためには、優良顧客の獲得は欠かせない課題である。実際に、本研究の対象事例である(株)良品計画では、顧客に対し、累計購買金額を基準とした会員ステージシステムを導入しており、ステージが上がる毎に特典を贈与することで、優良顧客への成長を促している。しかし、累計購買金額のみに着目した購買促進は、特に累計購買金額の少ない低ステージ顧客の購買意欲向上に対しても効果的なアプローチとは限らず、ステージランクや顧客ごとの購買傾向を反映した施策が望ましい。

一方、顧客の購買履歴データからその嗜好や購買傾向を把握する分析手法として潜在クラスモデルが提案されている。この手法は顧客の購買の背後に潜在クラスを仮定し、顧客や商品の多様性や異質性を表現することを可能とする。潜在クラスモデルの代表的な手法として Latent Dirichlet Allocation(LDA)[1]がある。LDAでは全顧客の購買履歴データから推定された商品間の共起性を基に各潜在クラスに商品の出現分布が与えられ、各顧客に潜在クラスへの所属確率の分布が仮定されるため顧客の購買嗜好が確率的に表現可能となる。

そこで本研究では、購買履歴データの分析に LDA を適用し、顧客のステージ向上に重要な商品を顧客ごと、ステージごとに抽出するための手法を提案する。この方法は、最上位ステージ以外の顧客が、それぞれ上位ステージに向上するための各商品の重要度を算出し、顧客ごと、またはステージごとに重要なアイテムを特定することを可能とする。さらに、(株)良品計画に提供頂いた無印良品のステージ情報付き購買履歴データに提案手法を適用し分析結果を示す。

## 2 Latent Dirichlet Allocation(LDA)

LDAを購買履歴データへ適用すると、顧客に潜在クラスの出現確率の分布、その潜在クラスごとに商品の出現確率の分布が仮定され、顧客数や商品数など次元数の膨大な情報を潜在クラスという軸で解釈できる。またモデルにベイズ統計の枠組みを導入しているので汎化性能の高いモデルとなる。

### 2.1 モデル式

顧客集合を  $\mathcal{D} = \{1, \dots, N\}$ 、商品集合を  $\mathcal{V} = \{1, \dots, V\}$  とし、 $K$  個の潜在クラス集合  $\mathcal{K} = \{1, \dots, K\}$  を仮定する。また、顧客  $d \in \mathcal{D}$  が潜在クラス  $k \in \mathcal{K}$  に所属する確率を  $\theta_{d,k}$ 、潜在クラス分布は  $\theta_d = (\theta_{d,1}, \dots, \theta_{d,K})^T$  とする。また、潜在クラス  $k$  の下で商品  $v \in \mathcal{V}$  が出現する確率を  $\phi_{k,v}$  とし、潜在クラス  $k$  の商品の出現確率の分布(商品分布)を  $\phi_k = (\phi_{k,1}, \dots, \phi_{k,V})^T$  と表記する。顧客ごとの潜在クラス分布  $\theta_d$ 、潜在クラスごとの商品分布  $\phi_k$  にはそれぞれ  $\alpha = (\alpha_k)$ 、 $\beta = (\beta_v)$  をパラメータとするディレクレ事前分布を仮定する。以上の下で、顧客  $d$  が商品  $v$  を購買する確率  $P(v|d)$  は式(1)で表現される。

$$P(v|d) = \sum_{k=1}^K \int \theta_{d,k} P(\theta_{d,k} | \alpha_k) \phi_{k,v} P(\phi_{k,v} | \beta_v) d\theta_{d,k} d\phi_{k,v} \quad (1)$$

このモデルのパラメータを購買履歴データにより学習し推定することで顧客の購買傾向を把握することができる。本研究

では購買履歴データに LDA を適用し、得られた潜在クラス分布と商品分布を活用した分析手法を提案する。

## 3 提案手法

### 3.1 概要

本研究では顧客のステージ向上に寄与するという意味で重要な商品を抽出することを目的とする。ここで事前分析により、累計購買金額によって購買嗜好に差があることが明らかになっている。累計金額によってステージが決定される状況下で、低ステージ顧客が高ステージ顧客の購買嗜好に推移すれば、低ステージ顧客のステージ向上が期待できると言える。しかし、同ステージ内には様々な購買嗜好の顧客が存在するため、低ステージ顧客に対して一様に高ステージ顧客の購買嗜好に推移させることは難しい。そこで、対象顧客の購買嗜好を大きくは変えずに、高ステージの購買嗜好へ促せれば購買嗜好の推移がしやすいと考えられる。本研究では各ステージの顧客に対し、それぞれの購買嗜好に合った高ステージ顧客の購買嗜好に推移するために重要となる商品を抽出するという着想のもとで提案手法を考案する。

以降は、重要商品を顧客ごと、またステージごとに抽出する方法について述べる。顧客ごとに抽出する方法では、顧客  $d$  の購買嗜好と顧客  $d$  の目指すべき高ステージの購買嗜好を算出し、これに基づく各商品の重要度スコアを算出することにより重要商品を算出する。ステージごとに抽出する方法では、顧客ごとの方法と同様のアルゴリズムで、ステージごとの各商品の重要度スコアを算出することにより重要商品を算出する。本手法はステージ向上を目的とするため、最上位ステージの顧客、最上位ステージの重要商品は抽出しない。

### 3.2 顧客ごとの重要商品抽出

本節では、顧客ごとにステージ向上の意味で重要な商品抽出法について述べる。まず顧客  $d$  の購買嗜好を LDA の適用により、潜在クラス分布  $\theta_d$  として抽出する。顧客  $d$  がステージを向上しやすく、かつ推移しやすい購買嗜好を顧客  $d$  の目標とする潜在クラス分布  $\theta'_d$  として式(2)で算出する。

$$\theta'_d = \frac{1}{h} \sum_{l \in D_d} \theta_l \quad (2)$$

式(2)において  $D_d$  は、類似した購買嗜好には推移しやすいという着想のもと、顧客  $d$  の潜在クラス分布との KL 情報量の意味で最も近い(顧客  $d$  の近傍の)潜在クラス分布をもつ顧客  $d$  よりも上位のステージの顧客  $h$  人の集合とした。

顧客  $d$  が、目標とする潜在クラス分布  $\theta'_d$  に推移するために、増加及び減少すべき購買嗜好を把握する必要がある。そのため各要素が、各潜在クラスの重要度を表す潜在クラスウエイトベクトル  $\delta_d = (\delta_{d,1}, \dots, \delta_{d,K})^T$  を式(3)で算出する。

$$\delta_d = \theta'_d - \theta_d \quad (3)$$

顧客  $d$  の潜在クラスウエイトベクトル  $\delta_d$  に基づき、顧客  $d$  の商品  $v$  の重要度スコア  $\gamma_{d,v}$  を式(4)で算出する。

$$\gamma_{d,v} = \sum_k \delta_{d,k} \phi_{k,v} \quad (4)$$

式(4)は顧客  $d$  の各潜在クラスに対する重要度  $\delta_{d,k}$  を重みとして、各潜在クラス  $k$  の商品  $v$  の出現確率  $\phi_{k,v}$  を足し合わせている。スコア上位の商品が顧客  $d$  の重要商品となる。以上をまとめると、顧客ごとの重要商品算出のアルゴリズムは以下のようなになる。

- Step1) LDA を用い顧客  $d$  の潜在クラス分布  $\theta_d$ , 潜在クラスごとの商品分布  $\phi_k$  を推定する.
- Step2) 顧客  $d$  の目標とする潜在クラス分布  $\theta'_d$  を式 (2) より算出する.
- Step3) 顧客  $d$  の潜在クラスウェイトベクトル  $\delta_d$  を式 (3) より算出する.
- Step4) 顧客  $d$  における商品  $v$  の重要度スコア  $\gamma_{d,v}$  を式 (4) より算出する.

### 3.3 ステージごとの重要商品抽出

本節では、ステージごとの重要商品抽出方法について述べる。各ステージの平均的嗜好を各ステージの代表顧客の嗜好と考え、3.2 節の顧客ごとの重要商品算出のアルゴリズムに適用する方法である。アルゴリズムにおいて顧客  $d$  を対象ステージの代表顧客とし適用する。各ステージの平均的嗜好は各顧客をステージ情報により層別し、ステージごとに顧客の潜在クラス分布の平均値を求めることにより得る。アルゴリズムの Step1 では、ステージの代表顧客がステージの平均的嗜好を持つと考える。Step2 では、式 (2) において  $D_d$  を対象ステージの代表顧客近傍の、上位ステージ顧客  $h$  人の集合として、対象ステージの代表顧客の目標とする潜在クラス分布を算出する。Step3,4 で対象ステージの代表顧客についてそれぞれ式 (3), (4) を用いることで、対象ステージの各商品の重要度スコアを算出することができる。

## 4 実データを用いた分析

### 4.1 分析条件

提案手法を用いて無印良品の 1 年間の購買履歴データを分析する。無印良品では会員ステージが低ステージ 0 から高ステージ 4 までの 5 段階に分類される。本分析では、年間累計商品購買個数が 5 個以上 99 個以下の顧客からランダムに抽出した 10,000 人を分析対象とした。データには中分類と小分類の情報が付与されている。小分類では商品名を、中分類では商品の属するカテゴリを把握することができる。商品については小分類である 21,450 アイテムを用いた ( $V = 21,450$ )。LDA のパラメータ推定は周辺化ギブスサンプリング法を適用した。事前実験より潜在クラス数  $K = 10$ 、ディレクレ事前分布のパラメータ  $\alpha_k = 0.05$ ,  $\beta_v = \frac{1}{V}$ 、ギブスサンプリングの回数は 1,000 回に設定した。ここでは、ステージごとに着目した分析結果を解釈することにする。

### 4.2 各ステージの平均の潜在クラス分布

ステージごとの平均の潜在クラス分布を図 1 に示す。

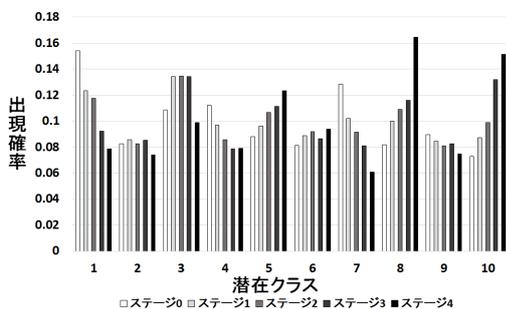


図 1. 各ステージの平均の潜在クラス分布

図 1 よりステージごとに購買嗜好に差があることがわかる。ここでステージにより出現確率に変化が見られた潜在クラスに着目してみる。低ステージから高ステージにかけ出現確率が著しく低下した潜在クラスは 1, 4, 7 であり、これらの潜在クラスに共通する特徴に注目すると、食品などの小物の商品が多く購買されていることがわかった。一方、中分類のカテゴリに着目するとこれらの各潜在クラス内には出現確

率の高い 2, 3 個の少数のカテゴリが多くを占めているという特徴が明らかになった。一方、低ステージから高ステージにかけて著しく増加した潜在クラスは 5, 8, 10 であり、これらの潜在クラスに共通する特徴は、大型の商品が出現しやすいこと、中分類のカテゴリに着目すると潜在クラス内には突出して高い出現確率のカテゴリはなく、多様なカテゴリが出現しやすいことである。このような結果から、低ステージ顧客は小物の商品や特定のカテゴリに属する商品を集中的に購買する傾向があるのに対し、高ステージ顧客は大型の商品を購買しやすく、かつ多様なカテゴリの商品に対して購買意欲が高いと考えられる。

### 4.3 各ステージの重要潜在クラス及び重要商品

各ステージの潜在クラスウェイトベクトル、商品の重要度スコアの上位 5 件の商品を重要商品として以下に示す。ここで、潜在クラス分布近傍  $D_d$  の人数  $h$  は 20 人とした。

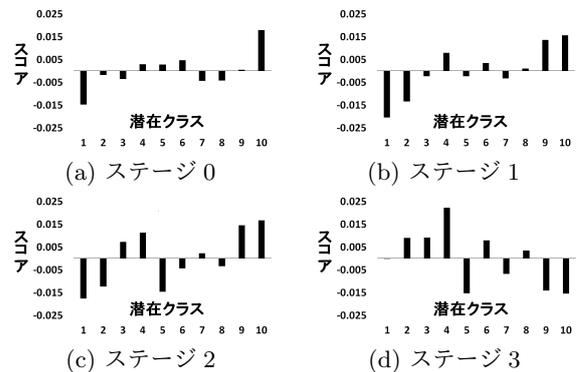


図 2. ステージごとの平均潜在クラスウェイトベクトル

図 2 より、ステージ 2 までは重要潜在クラスは 9, 10 であるが、ステージ 3 では 4 である。これよりステージ間で重要な潜在クラスが異なることがわかる。次にステージごとに算出した重要な商品を表 1 に示す。

表 1. ステージごとの重要商品 (単位:  $10^{-4}$ )

順位	ステージ 0		ステージ 1		ステージ 2		ステージ 3	
	スコア	商品	スコア	商品	スコア	商品	スコア	商品
1	1.9	生活雑貨 A	2.8	生活雑貨 F	3.0	生活雑貨 F	3.8	衣服雑貨 C
2	1.8	生活雑貨 B	2.4	生活雑貨 G	2.9	生活雑貨 G	3.7	生活雑貨 E
3	1.5	生活雑貨 C	2.2	生活雑貨 A	2.4	衣服雑貨 A	3.5	生活雑貨 J
4	1.5	生活雑貨 D	1.8	生活雑貨 H	2.2	生活雑貨 H	3.2	生活雑貨 I
5	1.4	生活雑貨 E	1.8	生活雑貨 I	2.1	衣服雑貨 B	3.2	生活雑貨 K

表 1 より、提案手法により抽出された商品を分析すると、低ステージには比較的大型の商品がステージ向上のために重要とされることがわかる。一方、高ステージでは衣類や小型の生活雑貨などの商品が重要であると判断されている。多くの高ステージ顧客はすでに大型の商品を購買する傾向にあるが、低ステージでは購買されていない小物商品が重要とされ、これらの商品の販売促進の効果があると解釈できる。

## 5 まとめと今後の課題

本研究では、購買履歴データに対して、LDA を適用することで、各顧客の潜在的な購買傾向を抽出し、購買傾向とステージ情報を考慮した顧客のステージ向上のための重要商品を抽出する手法を提案した。この方法は、抽出した重要商品から顧客のステージ向上を促す施策として、Web 上での顧客ごとの商品推薦や、クーポン発行による販売促進などに活用されるものと考えられる。今後の課題として、顧客が商品を購入した状況、顧客の属性、顧客の行動などの情報を考慮したステージ向上施策を考案すること等が挙げられる。

### 参考文献

- [1] Blei M.D., Ng Y.A., Jordan I.M., "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol.3, pp.993-1022 2003.