

生花 EC サイトを対象とした閲覧履歴に基づく購買行動分析に関する研究

1X17C046-0 鎌形祐志
指導教員 後藤正幸

1. 研究背景と目的

近年、EC サイトを通じた購買行動は一般的なものとなっており、生花商品もインターネットを通じて購入する消費者は日々増えている。また、消費者は EC サイト上の様々なページを閲覧し、商品の購入を検討することができる。ここで、顧客がサイトにアクセスしてから離脱するまでの一連の閲覧行動をセッション、その閲覧ページを閲覧系列と呼ぶ。一般に、全セッションのうち購入に至るものの割合は非常に低く、購買確率をいかに上げるかが重要な課題となっている。そこで、ユーザのニーズに適したサイト導線の設計など、購買意欲を高める効果的な施策が求められている。

ここで、本研究で対象となる生花 EC サイトは、主に贈答用の商品を取り扱っている。そのため、顧客は予め明確な購入用途のもとで、特集ページや金額ページなどから商品の絞り込みを行うことが多く、その閲覧順序はそれぞれの用途で異なると考えられる。例えば、“親族の誕生日”という用途で購入を行う顧客が特集ページから商品を絞り込む一方、受領者が取引先などであった場合には、金額から絞り込むことも考えられる。従って、最終的に購入される商品の特性とその絞り込み方の間には、関係性があると考えられ、これを分析することでユーザのニーズに適した商品をに辿り着きやすいサイト設計の一助となることが期待される。

一方、閲覧系列と購買の有無の関係性の分析手法として、松寄らのモデル [1] がある。この手法は、系列データに 1 次マルコフ性を仮定しその生起確率をモデル化する Latent Semantic Markov Chain[2](以下、LSMC) を拡張し、閲覧傾向と商品の購買有無の共起関係を分析可能な潜在クラスモデルである。これにより、購買の有無に伴う閲覧傾向の差異分析を可能とする。しかし、この研究では閲覧行動と購買有無の関係性に焦点が当てられており、購入される商品属性に依存する閲覧行動の違いを分析することはできない。一方、本研究では購買が行われたセッションのみを対象として、購買された商品属性と商品の絞り込みに伴う閲覧行動の関係性の分析を可能とするモデルを構築することが必要である。

そこで本研究では、松寄らのモデルをベースとし、生花 EC サイト上で購買が行われたセッションにおける、閲覧履歴データと購入商品の特性の分析に適用可能な拡張モデルを提案する。具体的には、商品の絞り込みを行うページの遷移を閲覧系列として扱い、LSMC で考慮する閲覧系列と、購入商品の特性の共起に基づく潜在クラスモデルの提案を行う。最後に、提案手法を実データに適用し、分析結果とその活用法について考察を行うことで、提案手法の有効性を示す。

2. 準備

2.1. 生花 EC サイトにおける閲覧行動の特徴

対象生花 EC サイトにおける購買行動の多くは、贈答を目的としている。顧客は自身の目的に合わせてサイト上のページ

を遷移しつつ、購入する商品を決定する。ここで、顧客が閲覧するページには、価格による商品絞り込みページや、イベント特集ページなどがある。

2.2. Latent Segment Markov Chain

LSMC とは、1 次マルコフ性を仮定した系列データに対して潜在クラスを導入したモデルであり、各潜在クラスのもとで、任意の系列が生起する確率を推定する。ここで、セッション集合を $\mathcal{S} = \{s_n : 1 \leq n \leq N\}$ 、 n 番目のセッション s_n においてユーザが閲覧したページ数を T_n とすれば、閲覧系列 \mathbf{x}_n は $\mathbf{x}_n = x_0^n, x_1^n, \dots, x_{T_n-1}^n$ で与えられる。ただし、 x_t^n はセッション s_n で t 番目に閲覧されたページを表す。

ここで、潜在クラス集合を $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$ と定義すると、ある潜在クラス z_k に所属するセッション s_n における閲覧系列 \mathbf{x}_n の生起確率は、以下の式 (1) のように表される。

$$P(\mathbf{x}_n | z_k) = P(x_0^n | z_k) \prod_{t=1}^{T_n-1} P(x_t^n | x_{t-1}^n, z_k) \quad (1)$$

なお、 $P(x_0^n | z_k)$ は初期分布であり、 $P(x_t^n | x_{t-1}^n, z_k)$ は z_k のもとで、 $t-1$ 番目にページ x_{t-1}^n を閲覧したのちに、 t 番目にページ x_t^n を閲覧する確率である。

2.3. 松寄らの研究

松寄らのモデルは LSMC を拡張し、閲覧系列と購買の有無の共起の背後に潜在クラスを仮定した潜在クラスモデルである。ここでセッション s_n における購買の有無を表す変数を w_n と定義する。ただし、購買を $w_n = 1$ 、非購買を $w_n = 0$ で表す。このとき、潜在クラス z_k の下でセッション s_n における閲覧系列 \mathbf{x}_n とその購買有無 w_n の生起確率は式 (2) で表される。

$$P(\mathbf{x}_n, w_n | z_k) = P(w_n | z_k) P(x_0^n | z_k) \prod_{t=1}^{T_n-1} P(x_t^n | x_{t-1}^n, z_k) \quad (2)$$

3. 提案手法

3.1. 着想

本研究で対象とする生花 EC サイト上における購買の多くは、贈答用の商品である。そのため、顧客の嗜好や顧客と受領者の関係性により、閲覧傾向と購入商品が異なると考えられる。そこで、本研究では購買に至ったセッションを対象として分析を行うことで、顧客の商品の絞り込みの傾向と購入商品の属性の関係性を明らかにすることを考える。具体的には、松寄らのモデルを拡張し、各セッションにおける LSMC でモデル化したページ遷移と購買商品の属性の組み合わせに対して、潜在クラスを仮定することで、それらの関係性を分析する。なお、本研究では商品属性として、生花商品の特徴づける色味や形状などの見た目および価格を用いており、選択した各商品属性毎にモデルを学習する。

3.2. 提案モデル

ページ遷移と商品属性に潜在クラスを仮定し、その共起関係をモデル化する提案手法の概要を図1に示す。

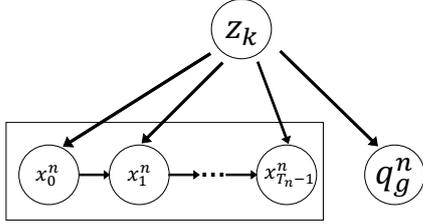


図 1: 提案手法のグラフィカルモデル

ただし、 q_g^n は n 番目のデータの $q_g \in \mathcal{Q}_g (1 \leq g \leq G)$ を表し、 \mathcal{Q}_g は第 g 商品属性の商品属性集合を表す。提案モデルの確率モデルを式 (3) に示す。

$$P(\mathbf{x}_n, q_g^n, z_k) = \sum_{k=1}^K P(z_k) P(q_g^n | z_k) P(x_0^n | z_k) \prod_{t=1}^{T_n-1} P(x_t^n | x_{t-1}^n, z_k) \quad (3)$$

このモデルの生起事象 (\mathbf{x}_n, q_g^n) は、閲覧系列が \mathbf{x}_n であるセッション s_n において、購入された商品の属性が q_g^n であることを意味する。ここで、 $P(\mathbf{x}_n | z_k)$ には LSMC を適用し、 $P(q_g^n | z_k)$ には多項分布を仮定する。

3.3. パラメータ推定アルゴリズム

閲覧系列と購入商品特性を持つ I 個の完全データについて、 i 番目のデータの生起確率は $P(\mathbf{x}_i, q_g^i, z_k)$ で表される。このとき、各パラメータは対数尤度関数を最大化するように EM アルゴリズムを用いて推定を行う。以下に対数尤度関数 LL を式 (4) に示す。

$$LL = \log \left(\prod_{i=1}^I \sum_{k=1}^K P(\mathbf{x}_i, q_g^i, z_k) \right) = \sum_{i=1}^I \log \sum_{k=1}^K P(z_k) P(q_g^i | z_k) P(x_0^i | z_k) \cdot \prod_{t=1}^{T_i-1} P(x_t^i | x_{t-1}^i, z_k) \quad (4)$$

4. 実データ分析

提案手法の有効性を検証するために、生花 EC サイト A における閲覧履歴データを使用した分析を行う。

4.1. 分析条件

本研究では、A 社提供の生花 EC サイト上のページ閲覧履歴データを対象とする。分析には、2019 年 3 月 1 日から 2020 年 2 月 28 日の閲覧データのうち、購買が起きた約 1 万件のセッションをサンプリングした。潜在クラス数 K は、解釈性の高いクラス数を探索的に求め $K = 3$ とした。また、分析に用いた商品属性とその区分を表 1 に示す。

表 1: 分析に用いた商品属性

色合い (q_1)	暖色系	寒色系	複色
価格 (円) (q_2)	0-4,999	5,000-6,999	7,000-
形状 (q_3)	アレンジメント	花束	その他

なお、提案手法においては、各商品属性の特徴を明確に抽出するため、用いる商品属性それぞれと閲覧系列との共起関係を分析している。

4.2. 分析結果と考察

提案手法による実データ分析について、商品属性として価格帯を用いた際の、潜在クラスごとの各価格帯の生起確率を表 2 に示す。また、各潜在クラスにおける、各ページから価格検索ページへの遷移確率を 3 に示す。

表 2: 各潜在クラスの価格帯生起確率

潜在クラス	z_1	z_2	z_3
0-4,999	0.81	0.40	0.02
5,000-6,999	0.13	0.41	0.58
7,000-	0.06	0.19	0.40

表 3: 各潜在クラスごとの遷移確率 (一部抜粋)

	z_1	z_2	z_3
お祝い→価格検索	0.025	0.031	0.042
季節性イベント→価格検索	0.009	0.008	0.015
お悔み→価格検索	0.010	0.021	0.210
価格検索→価格検索	0.532	0.553	0.640
花の特徴検索→価格検索	0.022	0.021	0.031

まず、表 2 から、 z_1 は低価格帯商品、 z_2 は低-中価格帯商品、 z_3 は中-高価格帯商品を購入するユーザと共起することが分かる。つまり、 z_1 から z_3 にかけて、購入商品の価格帯が徐々に高くなると言える。ここで、表 3 より、価格検索ページから価格検索ページに遷移する確率が z_1 から z_3 にかけて上昇していることがわかる。これは、低価格帯商品を購入するユーザは低予算で商品を探しているため、高価格帯商品をはじめから探す気がなく、高価格帯商品を購入するユーザの方が、様々な価格帯の商品で悩んでいることが考えられる。そのため、低価格帯商品の提示と同時に、少し高い価格帯の商品を提示することで、より高い価格の商品購買を促すことが可能になると考えられる。

5. まとめと今後の課題

本研究では、生花 EC サイトにおけるページ閲覧履歴と、購入商品属性の関係性を分析する手法を提案した。また、提案手法を実データに適用し、その有効性を示した。今後の課題として、適切なページ粒度を検討することが挙げられる。

謝辞

本研究を行うにあたり、貴重な実データを花キュービット株式会社よりご提供いただきました。深く感謝致します。

参考文献

- [1] 松崎祐樹, 三川健太, 後藤正幸, “マルコフ潜在クラスに基づく 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.44, no.12, pp.2034-2045, 2017.
- [2] 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.