

## 生花通信配達事業におけるイベント特性を考慮した顧客の購買行動分析モデル An Analytical Model of Customer Purchasing Behavior Considering Event Characteristics in Flower Delivery Business

北里礼\* 野中賢也\* 山下遥† 後藤正幸\*

Aya Kitazat Kenya Nonaka Haruka Yamashita Masayuki Goto

### 1 研究背景と目的

本研究では、インターネットを通じて生花を商材とした通信配達受注業務を行う生花ECサイト(以下、ECサイトA)を対象とし、顧客の特徴や購買行動の分析について検討する。生花は誕生日や母の日などの行事(以下、イベント)に紐づいて購入されるという特徴があり、イベントに着目した分析が有効である。生花商品の場合、顧客は商品のカテゴリや色味、価格などを参考に商品を選択・購入するが、顧客の属性(性別や年代)によって購買傾向が異なると考えられる。

このような複数の異なる特性をもつデータが混在した対象を分析するモデルとして、潜在クラスモデルが知られている[1]。このモデルによって、膨大な購買履歴データにおけるイベントと商品情報や顧客情報の背後に潜在的な変数を仮定することで、イベントと購買行動の関係性を潜在クラスという軸で解釈可能と考えられる。

そこで本研究ではイベント情報や顧客情報、ならびに商品情報間の関係性を表現する潜在クラスモデルを提案し、様々なイベントにおける顧客の特徴と購買行動を分析する手法を示す。また、提案モデルをECサイトAにおける実際の購買履歴データに適用して分析を行い、イベントごとの顧客属性や購買傾向を明らかにすることで、提案モデルの有用性を示す。

### 2 着想と提案モデル

本研究では、顧客の購買行動が様々な異なる嗜好を有した顧客の集合体から生成されていることから、複数の変数間の潜在的な関係のモデル化が可能な潜在クラスモデル Probabilistic Latent Semantic Analysis[1](以下、PLSA)の応用が分析目的に有効であると考えられる。具体的には、顧客の性別・年代や生花の用途(イベント情報)・色・単価・カテゴリの共起を潜在クラスによって紐づけ、分析を行う。従来のPLSAではアイテムと顧客を潜在クラスによって紐づけ、その共起関係を分析するが、本研究ではイベント情報とユーザ属性、商品情報(色・単価・カテゴリ・用途)を潜在クラスによ

て紐づけることで、イベントごとに購買される生花の特徴や顧客属性との関係について分析が可能となる。

いま、イベント集合を  $\mathcal{U} = \{u_i : 1 \leq i \leq I\}$ 、性別集合を  $\mathcal{E} = \{e_j : j = 1, 2\}$ 、年代集合を  $\mathcal{A} = \{a_l : 1 \leq l \leq L\}$ 、色集合を  $\mathcal{O} = \{o_c : 1 \leq c \leq C\}$ 、カテゴリ集合を  $\mathcal{G} = \{g_m : 1 \leq m \leq M\}$ 、単価を  $t$ 、 $K$  個の潜在クラス集合を  $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$  と定義する。このとき、提案する確率モデル式は式(1)で表される。

$$P(u_i, e_j, a_l, o_c, g_m, t, z_k) = P(u_i|z_k)P(e_j|z_k) \times P(a_l|z_k)P(o_c|z_k)P(g_m|z_k)P(t|z_k)P(z_k) \quad (1)$$

ただし、イベントや顧客の性別・年代、生花のカテゴリや色に関する確率分布  $P(u_i|z_k)$ 、 $P(e_j|z_k)$ 、 $P(a_l|z_k)$ 、 $P(o_c|z_k)$ 、 $P(g_m|z_k)$  には多項分布を仮定し、単価  $t$  には式(2)で表される正規分布を仮定する。

$$P(t|z_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left\{-\frac{(t-\mu_k)^2}{2\sigma_k^2}\right\} \quad (2)$$

また、 $\mu_k$ 、 $\sigma_k^2$  はそれぞれ潜在クラス  $z_k$  の正規分布の平均、分散を表している。

### 3 パラメータの学習

ECサイトAにおける全  $N$  件の購買履歴データのうち、 $n$  番目のイベントを  $b_n \in \mathcal{U}$ 、ユーザの性別を  $s_n \in \mathcal{E}$ 、年代を  $h_n \in \mathcal{A}$ 、生花の色を  $w_n \in \mathcal{O}$ 、カテゴリを  $r_n \in \mathcal{G}$ 、単価を  $p_n$  とすると、 $n$  番目の購買データはこれらの共起  $(b_n, s_n, h_n, w_n, r_n, p_n)$  で表現することができる。このとき、全  $N$  件の購買履歴データに対する対数尤度  $LL$  は以下の式(3)で表される。

$$LL = \sum_{n=1}^N \log \sum_{k=1}^K P(b_n, s_n, h_n, w_n, r_n, p_n, z_k) \quad (3)$$

提案モデルのパラメータはEMアルゴリズム[2]を用いて、対数尤度関数  $LL$  を局所最大化するように学習する。以下にEMアルゴリズムの更新式を示す。

E-step)

$$\gamma_{nk} = P(b_n, s_n, h_n, w_n, r_n, p_n | z_k) P(z_k) \quad (4)$$

$$P(z_k | b_n, s_n, h_n, w_n, r_n, p_n) = \frac{\gamma_{nk}}{\sum_{k=1}^K \gamma_{nk}} \quad (5)$$

\*早稲田大学

†上智大学

表 1: イベントと色の特徴

$z_k$	イベント	色	年代	$P(z_k)$
$z_2$	ご結婚記念日, 母の日, ご結婚祝い	ピンク, ミックス, ブルーとパープル	男性	0.041
$z_3$	お供え, 母の日	ミックス, ホワイト, ピンクとミックス系	女性	0.122
$z_6$	お供え, お盆, 遅れて母の日	ミックス, ホワイト, ピンク	男性, 女性	0.067
$z_{10}$	母の日, 敬老の日	ミックス, ピンク, 赤	男性	0.046
$z_{12}$	母の日	ミックス, ピンク, レッド	女性	0.217

M-step)

$$P(z_k) = \frac{\sum_{n=1}^N P(z_k|b_n, s_n, h_n, w_n, r_n, p_n)}{N} \quad (6)$$

$$P(u_i|z_k) = \frac{\sum_{n=1}^N \delta(b_n, u_i) P(z_k|b_n, s_n, h_n, w_n, r_n, p_n)}{NP(z_k)} \quad (7)$$

$$P(e_j|z_k) = \frac{\sum_{n=1}^N \delta(s_n, e_j) P(z_k|b_n, s_n, h_n, w_n, r_n, p_n)}{NP(z_k)} \quad (8)$$

$$P(a_l|z_k) = \frac{\sum_{n=1}^N \delta(h_n, a_l) P(z_k|b_n, s_n, h_n, w_n, r_n, p_n)}{NP(z_k)} \quad (9)$$

$$P(o_c|z_k) = \frac{\sum_{n=1}^N \delta(w_n, o_c) P(z_k|b_n, s_n, h_n, w_n, r_n, p_n)}{NP(z_k)} \quad (10)$$

$$P(g_m|z_k) = \frac{\sum_{n=1}^N \delta(r_n, g_m) P(z_k|b_n, s_n, h_n, w_n, r_n, p_n)}{NP(z_k)} \quad (11)$$

$$\mu_k = \frac{\sum_{n=1}^N P(z_k|b_n, s_n, h_n, w_n, r_n, p_n) p_n}{\sum_{n=1}^N P(z_k|b_n, s_n, h_n, w_n, r_n, p_n)} \quad (12)$$

$$\sigma_k^2 = \frac{\sum_{n=1}^N P(z_k|b_n, s_n, h_n, w_n, r_n, p_n) (p_n - \mu_k)^2}{\sum_{n=1}^N P(z_k|b_n, s_n, h_n, w_n, r_n, p_n)} \quad (13)$$

ここで,  $\delta(\alpha, \beta)$  は,  $\alpha = \beta$  であるときに 1, それ以外の場合に 0 を示すインジケータ関数である.

## 4 実データ分析

### 4.1 分析条件

実データ分析に用いるデータ収集期間は 2018 年 8 月 ~ 2019 年 9 月である. 全データ数  $N$  は 390,461 件となった. イベントの種類数は  $I = 40$ , 性別の種類数は  $J = 2$ , 年代の種類数は  $L = 12$ , 色の種類数は  $C = 45$ , カテゴリの種類数は  $M = 148$  である. さらに, 潜在クラスモデルにおけるクラス数は事前実験の結果, 解釈性の観点から  $K = 12$  とした.

### 4.2 分析結果と考察

提案モデルによって得られたイベントや生花の色, 性別などの  $z_k$  のもとでの生起確率と潜在クラスの生起確率  $P(z_k)$  を表 1 に示す. なお, 紙面の都合上, 議論対象のクラスや特徴量のみを抜粋して示す.

イベントと色は生起確率が 0.05 以上のものを降順で最大 3 つ目まで示した. また, 性別は原則上位のみ示

し,  $z_k$  のもとでの片方の性の生起確率  $P(e_j = 1|z_k)$  が 0.5 付近で, どちらの性別もほぼ等確率で生起しうる場合に両性を結果に示す.

$z_2, z_{10}, z_{12}$  はいずれも母の日が高い生起確率を有しているが,  $z_{10}, z_{12}$  で顧客層が女性かつ色味が暖色が好まれる傾向にあるのに対し,  $z_2$  では顧客層が男性で色味がブルーとパープルといった寒色の生花も高い生起確率となっている. ここから母の日というイベントにおいても購入者の性別によって好まれる花の色味が変わる可能性が示唆される.

また,  $z_3$  は母の日とお供えが高い生起確率を有し, 母の日とお供えの顧客の属性や購買傾向が類似するということを示唆する意外性のある結果となった. また,  $P(z_3)$  と  $P(z_{12})$  が特に高い確率となり, この 2 クラスのような購買傾向を持つ顧客が多いということがわかる. よってこれらのクラスに属する顧客に対する施策は重要と考えられる.

## 5 まとめと今後の課題

本研究では, イベントや花のカテゴリ, 色味や単価, 顧客の性別や年代の共起関係を表現した潜在クラスモデルを提案した. その結果, 各潜在クラスに関して様々な視点で分析することができ, 特に, 類似したイベントであっても性別による購買傾向の違いなどを把握することが可能となった. これらの結果を活用することで, イベントごとに性別や年代といったセグメントに対して異なる商品の推薦などができ, 購買数の増加につながる可能性があると期待できる.

今後の課題として, 提案モデルに取り込んだ変数以外にも例えばサイト閲覧履歴データなど様々な要素を加えることで, より詳細な分析を行うことが挙げられる.

## 参考文献

- [1] Thomas Hofmann, "Probabilistic latent semantic analysis", *Proceedings of the 15th Conference on Uncertainty in AI*, pp.289-296, (1999).
- [2] A. P. Dempster, et al, "Maximum likelihood from incomplete data via the em algorithm", *Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1, pp. 1-38, (1977).