

周期性とイベント効果に着目した消費者の購買行動分析モデルに関する研究

1X14C135-5 安井 一貴
指導教員 後藤 正幸

1 研究背景・目的

情報技術の発展に伴い、大量に蓄積されている消費者の購買履歴データを用いて消費者の購買行動を把握し、消費者の購買行動に合わせたマーケティング施策を講じることは、企業の売上向上のために有効である。消費者の購買行動には、それぞれの生活リズムに合わせて購買を行う周期的な購買行動や、祝日やクリスマスなどのイベント時に通常よりも多く購買する行動が含まれると考えられる。そのため、マーケティングにおいてこれらの周期的な購買行動やイベントへの反応度に着目し、消費者の購買行動を分析することは有用であると考えられる。消費者の周期的な購買行動とイベントへの反応度の両方を考慮することで、両者の関係性を把握することができるため、どちらか一方のみの分析よりも有用であることが期待される。

しかしながら、消費者の周期的な購買行動とイベントへの反応度の両方に着目した分析はほとんどなされていない。このような分析を可能とする手法として時系列データをトレンド関数、周期関数、イベント関数の3つに分解して表現する Prophet [1] がある。これを改良し、購買履歴データに対して適用することで、消費者の周期的な購買行動とイベントへの反応度の把握が期待される。

本研究では、消費者の一般的な行動と考えられる、購買行動の周期性とイベントへの反応度の両方に着目した消費者の購買行動分析モデルを提案する。まず、Prophet を用いて得られる消費者の購買行動の周期性とイベントに対する反応度から周期成分とイベント成分をそれぞれ抽出する。それをもとにデータ分布を複数の正規分布の混合により表現する手法である Gaussian Mixture Model (GMM) [2] を用いて、周期性とイベント反応度の双方による消費者のクラスタリングを行う。その後、周期成分のクラスタリング結果とイベント成分のクラスタリング結果を組み合わせることで分析を行う。これにより、消費者の周期性とイベントへの反応度の双方を考慮した購買行動分析が可能となる。本研究では、提案モデルを株式会社マクロミルから提供された実購買データに適用し、消費者の購買行動を分析することで、その有用性を示す。

2 準備

2.1 Prophet

時系列データにおいて、時刻 $t(1 \leq t \leq T)$ におけるデータを $y(t)$ としたとき、Prophet はこれをトレンド関数 $g(t)$ 、周期関数 $s(t)$ 、イベント関数 $h(t)$ の3項と誤差 ε_t に分解し、以下の式 (1) で表すモデルである。いま、 $g(t)$ 、 $s(t)$ 、 $h(t)$ はそれぞれ $y(t)$ の大局的な傾向、周期的な変動、祝日などのイベントへの反応度合を表す。

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (1)$$

以下では、本研究で扱う周期関数 $s(t)$ とイベント関数 $h(t)$ について述べる。

(a) 周期関数

周期関数は、時系列データの周期的変化を表した関数である。時刻 t での周期関数はフーリエ級数を用いて式 (2) で表現する。

$$s(t) = \sum_{m=1}^M \left(a_m \cos \left(\frac{2\pi mt}{P} \right) + b_m \sin \left(\frac{2\pi mt}{P} \right) \right) \quad (2)$$

いま、式 (2) において、 a_m 、 b_m は $m(1 \leq m \leq M)$ 番目の波におけるフーリエ係数、 P は周期係数である。

(b) イベント関数

イベント関数は、祝日などのイベントへの反応度を定量的に表現する関数である。イベント $\theta_l(1 \leq l \leq L)$ に対する反応度を κ_l とし、これが平均 0、分散 ν_l の正規分布に従うと仮定する。いま、 L 個のイベントへの反応度をベクトル

$\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_l, \dots, \kappa_L)^T$ とすると、イベント関数は式 (3) で表される。

$$h(t) = \mathbf{Z}(t)^T \boldsymbol{\kappa} \quad (3)$$

なお、 $\mathbf{Z}(t)$ は時刻 t のときイベント θ_l が行われていれば l 番目の要素が 1、それ以外の要素は 0 をとる L 次元ベクトルとする。

2.2 Gaussian Mixture Model

Gaussian Mixture Model (GMM) とは、データ \boldsymbol{x} を複数の正規分布の混合によって表現する統計モデルである。ここで、 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。GMM のモデル式は以下の式 (4) で表される。ただし、 $p(\boldsymbol{x}|z_k)$ は潜在クラス z_k のもとで \boldsymbol{x} が生起する確率で、平均 $\boldsymbol{\mu}_k$ 、分散共分散行列 $\boldsymbol{\Sigma}_k$ の多次元正規分布を表す。

$$p(\boldsymbol{x}) = \sum_{k=1}^K p(z_k) p(\boldsymbol{x}|z_k) \quad (4)$$

GMM は、潜在クラスごとに異なる多次元正規分布を仮定できるため、複雑な構造を持つデータのモデル化に有効であることが知られている。

3 提案モデル

3.1 概要

本研究では、消費者の購買履歴データから得られる周期的な購買傾向とイベントへの反応度の2つに着目してそれぞれ消費者のクラスタリングを行う。消費者の購買行動のモデル化のため、これらの結果を組み合わせることで、周期的な購買行動とイベントへの反応度の関係性を考慮した消費者の購買行動を分析できることが期待される。

提案モデルではまず、Prophet を用いて、各消費者の購買履歴データをトレンド関数、周期関数、イベント関数に分解する。このうち周期関数とイベント関数から、消費者の周期的な購買行動とイベントへの反応度の成分をそれぞれ抽出する。そして、抽出された各成分に対して、GMM を用いて学習し、モデル化を行う。GMM により購買傾向の周期性やイベントへの反応度で消費者のクラスタリングが可能となる。その後、各消費者を周期成分の潜在クラスへの所属確率、イベント成分の潜在クラスへの所属確率がそれぞれ最大になるクラスに所属すると仮定し、クロス分析を行う。具体的には、周期成分の各潜在クラスに所属する消費者のうち、イベント成分の各潜在クラスに所属する消費者の割合を求め、同様に、イベント成分の各潜在クラスに所属する消費者のうち、周期成分の各潜在クラスに所属する消費者の割合を求め、購買傾向の周期性とイベントへの反応度の関係を分析する。

3.2 各成分における購買行動のクラスタリング

いま、消費者集合を $\mathcal{U} = \{u_i: 1 \leq i \leq I\}$ とする。まず、消費者 u_i の購買履歴データに対して Prophet を用いて、あるアイテムにおける購買数量をトレンド関数、周期関数、イベント関数に分解する。ここで、周期を D とし、消費者 u_i の周期関数 $s(t)$ から 1 周期の周期効果 (D 次元) を取り出す。 D 次元のうち、 d 期目 ($1 \leq d \leq D$) の周期効果を s_{id} とし、周期成分を $\mathbf{s}_i = (s_{i1}, \dots, s_{id}, \dots, s_{iD})^T$ と表す。同様に、イベント関数 $h(t)$ から全 E 個のイベント効果を取り出す。そのうちの e 番目 ($1 \leq e \leq E$) のイベント効果を h_{ie} とし、イベント成分を $\mathbf{h}_i = (h_{i1}, \dots, h_{ie}, \dots, h_{iE})^T$ と表す。

次に、抽出された周期成分 \mathbf{s}_i とイベント成分 \mathbf{h}_i に対して GMM を適用し、消費者のクラスタリングを行う。ここで、周期成分に対しては、 J 個の潜在クラス集合 $\mathcal{V} = \{v_j: 1 \leq j \leq J\}$ を、イベント成分に対しては、 L 個の潜在クラス集合 $\mathcal{W} = \{w_l: 1 \leq l \leq L\}$ をそれぞれ仮定する。周期成分 \mathbf{s}_i とイベント成分 \mathbf{h}_i の GMM モデル式は以下の式 (5)、(6) のように表すことができる。ここで、 $p(v_j)$ 、 $p(w_l)$ は各

潜在クラス v_j, w_l の生起確率, $p(\mathbf{s}_i|v_j)$ は平均 $\boldsymbol{\mu}_j$, 分散共分散行列 $\boldsymbol{\Omega}_j$ の多次元正規分布, $p(\mathbf{h}_i|w_l)$ は平均 $\boldsymbol{\nu}_l$, 分散共分散行列 $\boldsymbol{\Gamma}_l$ の多次元正規分布を表すものとする.

$$p(\mathbf{s}_i) = \sum_{j=1}^J p(v_j)p(\mathbf{s}_i|v_j) \quad (5)$$

$$p(\mathbf{h}_i) = \sum_{l=1}^L p(w_l)p(\mathbf{h}_i|w_l) \quad (6)$$

3.3 両成分を考慮した購買行動分析モデルの作成

学習した GMM を用い, 消費者 u_i は所属確率 $P(v_j|u_i)$ と $P(w_l|u_i)$ がそれぞれ最大となる潜在クラスに所属すると仮定することで, クロス分析が可能となる. 周期成分の潜在クラス v_j に所属する消費者のうち, イベント成分の潜在クラス w_l に所属する消費者の割合を求めることで, 周期成分からイベント成分への関係の強さが表現できる. また, 同様にイベント成分から周期成分への割合を求めることでイベント成分から周期成分への関係の強さを表現できる. これにより, ある周期的な購買行動をとる消費者はある特定のイベントに反応しやすいといった, 購買行動の周期性とイベントへの反応度の関係を表現した購買行動分析が可能となる.

周期成分の潜在クラス v_j のうち, イベント成分の潜在クラス w_l にも所属する消費者の割合を α_{jl} , イベント成分の潜在クラス w_l のうち, 周期成分の潜在クラス v_j にも所属する消費者の割合を β_{lj} とし, 以下の式 (7), (8) で表す. ここで, \mathcal{A}_j はクラス v_j に所属する消費者集合, \mathcal{B}_l はクラス w_l に所属する消費者集合とする.

$$\alpha_{jl} = \frac{n(\mathcal{A}_j \cap \mathcal{B}_l)}{n(\mathcal{A}_j)} \quad (7)$$

$$\beta_{lj} = \frac{n(\mathcal{B}_l \cap \mathcal{A}_j)}{n(\mathcal{B}_l)} \quad (8)$$

なお, $n(\mathcal{X})$ は集合 \mathcal{X} に所属する消費者の人数を表す. ここで, v_j と w_l のすべての組み合わせにおいて α_{jl}, β_{lj} を算出することで, 消費者の周期性とイベントへの反応度の双方を考慮した購買行動分析が可能となる. 具体的には, 上式の中で割合が高くなった j と l の組み合わせは v_j, w_l 間に同じ消費者が多く存在するため, 強い関係を持つと考えられる.

4 実験

提案モデルの有効性を示すために, 実データを用いた検証を以下に示す. 実験に用いたデータは株式会社マクロミル提供の消費者購買履歴データ Q P R で, 2015 年の 1 月 1 日から 2015 年の 12 月 31 日までの期間で総購買数の多い菓子カテゴリの購買履歴データを対象とした. 対象ユーザ数は $I = 7,763$ である. ここで, Prophet における周期成分として週周期 ($P = 7, D = 7$) に, イベント成分として, 一般的な日本のイベントと考えられる年始, バレンタインデー, 5 月の連休 (GW), お盆, ハロウィン, クリスマス, 年末の 7 種類 ($E = 7$) に着目した. また, 事前実験より周期関数における波の個数を $M = 3$, GMM における各成分の潜在クラス数をそれぞれ $J = 7, L = 7$ と設定した.

4.1 実験結果

消費者 u_i の周期成分とイベント成分に対して GMM を用いて消費者の分類を行い, 得られた潜在クラス v_j, w_l に解釈を与えた結果を以下の表 1, 2 に示す.

表 1. 各潜在クラスの特徴 (周期成分)

j	所属している消費者の特徴
1	平日に購買する
2	土・日に購買する
3	月・水に購買する
4	火・木・金に購買する
5	土・日に購買する
6	平日に購買する
7	日に購買する

表 2. 各潜在クラスの特徴 (イベント成分)

l	所属している消費者の特徴
1	クリスマス
2	年始, 年末
3	年始, GW, 年末
4	年始, GW, お盆
5	バレンタイン, GW, お盆, ハロウィン, クリスマス
6	年始
7	バレンタイン, お盆, ハロウィン, クリスマス

表 1 より, 周期成分に GMM を適用することにより, 消費者を周期的な購買傾向に基づきクラスターリングできていることが分かる. また, 表 2 より, イベント成分に GMM を適用することにより, 反応するイベントの種類によって消費者をクラスターリングできていることも分かる.

次に, 周期成分とイベント成分の各潜在クラスの双方への所属割合 α_{jl}, β_{lj} を表 3, 4 に示す. また, 各行で最大の割合を太字で示している.

表 3. 周期成分からイベント成分への所属割合 α_{jl}

$j \setminus l$	1	2	3	4	5	6	7
1	0.466	0.181	0.011	0.025	0.158	0.151	0.009
2	0.042	0.052	0.166	0.000	0.460	0.000	0.280
3	0.236	0.183	0.053	0.002	0.446	0.025	0.055
4	0.178	0.163	0.053	0.002	0.516	0.012	0.075
5	0.192	0.150	0.093	0.002	0.487	0.004	0.073
6	0.212	0.070	0.001	0.159	0.019	0.537	0.002
7	0.416	0.200	0.024	0.008	0.263	0.070	0.019

表 4. イベント成分から周期成分への所属割合 β_{lj}

$l \setminus j$	1	2	3	4	5	6	7
1	0.516	0.005	0.107	0.069	0.043	0.107	0.153
2	0.405	0.012	0.168	0.128	0.067	0.072	0.148
3	0.115	0.183	0.225	0.195	0.195	0.004	0.084
4	0.248	0.000	0.007	0.007	0.004	0.709	0.025
5	0.206	0.064	0.239	0.237	0.128	0.012	0.114
6	0.347	0.000	0.024	0.009	0.002	0.565	0.053
7	0.081	0.273	0.205	0.242	0.135	0.007	0.057

表 3 の α_{45} に着目すると, 火・木・金曜日にアイテムを多く購買する消費者群は年末年始以外のイベントに反応していることが分かる. 同様に表 4 の β_{46} の値に着目すると, 年始, GW, お盆にアイテムを多く購買する消費者群は平日にアイテムを購買することが分かる. これにより, アイテムをある周期で購買する消費者がどのようなイベントに反応するのか, またその逆の関係を分析することができる.

以上より, 提案モデルを用いて消費者の購買傾向の周期性とイベントへの反応度を組み合わせ, これらの関係性という観点から消費者の購買傾向を分析できることが示された.

5 まとめと今後の課題

本研究では, 消費者の購買履歴データから得られる購買傾向の周期性とイベントへの反応度の両方を組み合わせた購買行動分析モデルを提案した. また, 実データに提案したモデルを適用することで, 周期的な購買傾向とイベントへの反応度の両方を考慮した消費者の購買行動を表現するモデルとして有効なモデルであることを示した.

今後の課題として, 消費者の周期的な購買傾向が途中で変化した場合の消費者特性の取り込み方の検討や, イベントに対して当日, 数日前などいつ反応するのかといったイベントへの反応速度を考慮した消費者の購買行動分析モデルの作成などが挙げられる.

参考文献

- [1] S. J. Taylor, B. Letham, "Forecasting at scale," *PeerJ Preprints*, Vol. 5, pp. 1–25, 2017.
- [2] Z. Zhang, B. T. Dai, A. K. H. Tung, "Estimating local optimums in EM algorithm over Gaussian mixture model," *Proc. 25th International Conference on Machine Learning*, pp. 1240–1247, 2008.