

潜在クラスモデルに基づく初期購買を考慮したRFM分析モデルに関する一考察 Analysis of Purchase History Data Using Extended RFM Analysis Based on Latent Class Model

張 倩[†]
Qian Zhang

山下 遥[†]
Haruka Yamashita

三川 健太[§]
Kenta Mikawa

後藤 正幸[†]
Masayuki Goto

1 研究背景・目的

顧客の購買データを解析する際に、その特徴をよく表す3つの指標を用いた方法として、RFM (R:最終購買日, F:購買頻度, M:購買金額) 分析が幅広く活用されている。一方、著者らはRFM分析の指標を用いて特徴的な顧客の分析を行うために、潜在クラスモデルに基づくRFMモデルを提案している [1]。これにより、潜在的に類似した嗜好を持つ顧客群をクラスタリングすると同時に、潜在クラスの特徴を分析することが可能となる。しかし、その顧客がいつから購買を始めたかという情報は、顧客の入れ替わりのある小売店では重要であるもののモデルには取り入れられていない。また、潜在クラスモデルによる分析結果を分かり易く可視化する方法が望まれる。本研究では、初期購買日を新たな特徴量とした潜在クラスモデルに基づくRFMモデルを提案する。さらに、拡張したモデルを活用した自己組織化マップ (SOM) により、分析結果を可視化する方法を示す。

2 潜在クラスモデルに基づくRFMモデル [1]

購買行動のという特徴量の背後にR, F, Mに注目するRFM分析に、嗜好の類似したユーザの集合である潜在クラスを仮定し、これらの共起関係を潜在クラスによる条件付確率分布で表した潜在クラスモデルが提案されている [1]。

いま、RFMのそれぞれの特徴量を x_{ni} ($i \in \{1, \dots, L\}$, $n = \{1, 2, 3\}$)、潜在クラスを z_k ($k \in \{1, \dots, K\}$)、 L を総ユーザ数、 K を総潜在クラス数とする。また、 $n=1$ はR, $n=2$ はF, $n=3$ はMの特徴量とする。このとき、潜在クラスにおける各特徴量R, F, Mの出現傾向を表す確率密度関数 $P(x_{ni}|z_k)$ を用いて、RFMに基づく顧客の購買行動を表す確率分布 $P(x_{1i}, x_{2i}, x_{3i})$ を式 (1) で定式化する。

$$P(x_{1i}, x_{2i}, x_{3i}) = \sum_{k=1}^K P(z_k) P(x_{1i}|z_k) P(x_{2i}|z_k) P(x_{3i}|z_k) \quad (1)$$

ただし、式 (1) における $P(x_{ni}|z_k)$ はそれぞれ独立に正規分布に従うものと仮定し、その平均を μ_{nk} 、分散を σ_{nk}^2 とする。このときのグラフィカルモデルを図1に示す。

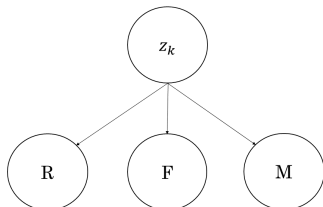


図1. 潜在クラスモデルに基づくRFMモデル [1]

$P(x_{ni}|z_k)$ は式 (2) により表現することができる。

$$P(x_{ni}|z_k) = \frac{1}{\sqrt{2\pi\sigma_{nk}^2}} \exp \left\{ -\frac{(x_{ni} - \mu_{nk})^2}{2\sigma_{nk}^2} \right\} \quad (2)$$

これらのパラメータ μ_{nk} , σ_{nk}^2 , $P(z_k)$ はEMアルゴリズムにより推定する。

[†]早稲田大学

[§]湘南工科大学

3 初期購買を考慮したRFM分析モデル

前述の潜在クラスを用いたRFM分析では、最終購買日、購買頻度、購買金額に着目していた。しかし例えば、ある一年間のR, F, Mの値が同じ顧客を考えた場合、このとき、最初の購買日 (初期購買日) の値により、顧客の解釈が大きく変化する。そこで本研究では、初期購買日を新たな特徴量として導入したモデルの構築を行う。

3.1 初期購買を考慮したRFM分析モデルの定式化

いま、 i 番目の顧客の初期購買日を x_{4i} とする。潜在クラスにおける最初購買日の出現確率 $P(x_{4i}|z_k)$ を用いて、初期購買日を考慮したモデルを式 (3) で定式化する。

$$P(x_{1i}, x_{2i}, x_{3i}, x_{4i}) = \sum_{k=1}^K P(z_k) P(x_{1i}|z_k) P(x_{2i}|z_k) P(x_{3i}|z_k) P(x_{4i}|z_k) \quad (3)$$

ただし、 $P(x_{4i}|z_k)$ は、式 (1) の $P(x_{ni}|z_k)$ と同様に正規分布に従うものと仮定する。

3.2 パラメータの推定

式 (3) のパラメータ $P(z_k)$, μ_{nk} , σ_{nk}^2 を、潜在クラスに基づくRFM分析 [1] と同様に、E-Step, M-Step を繰り返すことで尤度を最大にするパラメータを推定するEMアルゴリズムによって推定する。提案モデルに対して構成したEMアルゴリズムのパラメータ推定式は以下のように表される。

[E-Step]

$$P(z_k|x_{1i}, x_{2i}, x_{3i}, x_{4i}) = \frac{P(z_k) P(x_{1i}|z_k) P(x_{2i}|z_k) P(x_{3i}|z_k) P(x_{4i}|z_k)}{\sum_k P(z_k) P(x_{1i}|z_k) P(x_{2i}|z_k) P(x_{3i}|z_k) P(x_{4i}|z_k)} \quad (4)$$

[M-Step]

$$P(z_k) = \frac{1}{L} \sum_i P(z_k|x_{1i}, x_{2i}, x_{3i}, x_{4i}) \quad (5)$$

$$\mu_{nk} = \frac{\sum_i x_{ni} P(z_k|x_{1i}, x_{2i}, x_{3i}, x_{4i})}{\sum_i P(z_k|x_{1i}, x_{2i}, x_{3i}, x_{4i})} \quad (6)$$

$$\sigma_{nk}^2 = \frac{\sum_i (x_{ni} - \mu_{nk})^2 P(z_k|x_{1i}, x_{2i}, x_{3i}, x_{4i})}{\sum_i P(z_k|x_{1i}, x_{2i}, x_{3i}, x_{4i})} \quad (7)$$

4 実データ解析

提案手法の有効性を検討するため、本研究では、経営科学系研究部会連合協議会主催、平成27年度データ解析コンペティションで提供された、株式会社アイディーズのi-codeデータサービスによる、某スーパーマーケットの代表的店舗における2014年1月から2014年12月の1年間の購買履歴データを用いて解析する。顧客数は $L=111,753$ であり、潜在クラス数 $K=20$ と設定した。初期購買を考慮した提案モデルと従来モデル [1] での分析の結果を表1, 表2に示す。

表1. 提案モデルによる分析結果

	初期購買日	R	F	M	割合
z_1	183.7	183.7	1.0	1267.0	33.64%
z_2	230.1	135.4	2.0	2584.9	14.57%
z_3	254.5	107.6	3.0	3969.8	8.47%
z_4	215.5	44.7	4.0	5927.0	2.91%
z_5	323.2	138.2	4.0	4854.8	2.89%
z_6	282.6	80.1	5.0	6909.6	4.21%
z_7	280.3	52.8	6.5	6511.8	2.47%
z_8	339.4	148.7	7.6	8246.3	2.40%
z_9	226.3	13.9	8.8	10423.7	3.09%
z_{10}	349.0	36.3	10.1	11674.2	3.28%
z_{11}	280.1	77.1	12.5	18488.0	2.33%
z_{12}	340.7	16.0	16.0	21301.7	3.91%
z_{13}	357.3	136.7	17.8	22693.6	2.17%
z_{14}	222.2	8.2	21.3	28605.4	1.91%
z_{15}	354.8	11.2	31.9	44271.8	4.29%
z_{16}	360.9	102.4	52.2	71734.4	0.90%
z_{17}	278.2	4.8	57.2	82986.8	1.06%
z_{18}	359.7	5.6	74.2	108172.0	3.31%
z_{19}	313.2	51.9	96.6	159470.0	0.62%
z_{20}	362.5	2.4	182.1	314337.0	1.55%

表2. 従来手法 [1] による分析結果

	R	F	M	割合
z_1	183.8	1.0	1261.5	33.62%
z_2	135.4	2.0	2583.5	14.56%
z_3	107.6	3.0	3968.8	8.47%
z_4	91.4	4.0	5370.1	5.79%
z_5	80.1	5.0	6909.7	4.21%
z_6	73.0	6.0	8371.8	3.10%
z_7	102.9	7.3	6658.9	1.53%
z_8	9.7	8.4	8766.3	1.79%
z_9	68.3	9.2	10771.2	2.32%
z_{10}	24.9	11.2	12903.6	3.05%
z_{11}	203.3	12.3	14676.9	1.54%
z_{12}	8.2	15.6	19163.1	3.70%
z_{13}	53.2	16.0	22780.8	2.50%
z_{14}	19.6	25.5	36713.3	2.78%
z_{15}	4.6	31.6	40820.5	2.89%
z_{16}	128.8	36.3	48021.0	1.21%
z_{17}	8.1	55.3	80366.0	3.21%
z_{18}	73.0	104.2	163793.0	0.46%
z_{19}	3.8	107.5	165763.0	2.37%
z_{20}	1.8	219.7	397618.0	0.89%

表1, 表2の各列はデータの一年間の初期購買日, 最終購買日(数字が小さい方が現時点に近い), 総購買頻度, 総購買金額, 各潜在クラスの顧客割合を表している。

表1と表2は, 提案モデルと従来モデルにより得られた各潜在クラスをFの値の昇順に並べたものである。例えば, 表1および表2の潜在クラス z_1, z_2 では, Rの値が大きく, FとMの値が小さいことから, 非優良顧客のクラスと解釈することができる。一方, 表1の z_{20} , 表2の z_{19}, z_{20} は, Rの値が小さく, FとMの値が大きいことから, 優良顧客と解釈可能である。また, 表1の z_4, z_5 に着目すると, Fの値が4.0で等しくなっているのに対して, それ以外の指標の値は大きく異なっている(z_4 の方が初期購買日とRが小さく, Mが大きいため)ことから, z_4 は通常顧客, z_5 は離反顧客と解釈することができる。ここで, 表2の z_4 に着目すると, 表1における z_4 と z_5 の中間的な値をとっている。初期購買日の指標を追加したことにより, 離反顧客のクラスを作ることができたものと思われる。

5 SOMによる潜在クラスの分析

自己組織化マップ(SOM: Self-Organizing Map) [2]は, T.Kohonenにより提案された教師なしのニューラルネットワークモデルで, 高次元の属性を持つデータを非線形写像により2次元平面上へデータ間の類似性に応じてプロットする手法である。

本研究では, 初期購買日, R, F, Mの4次元の特徴量を用い, SOMにより2次元平面上へ顧客をプロットし, 顧客

の類似性を可視化する。初期購買を考慮したモデルの結果を用い, 各顧客の各潜在クラスへの所属確率に着目し, その値が最も大きいクラスに顧客を割り当てる。さらに, クラスごとの顧客の初期購買日, R, F, MのデータをSOMに当てはめて顧客の関係性を可視化する。

ここでは, 提案モデルによって求めた潜在クラスのうち, 非優良顧客を表している潜在クラス z_1 と優良顧客を表している潜在クラス z_{18} を例とし, 潜在クラスの可視化を行う。SOMによる可視化の結果を図2と図3に示す。

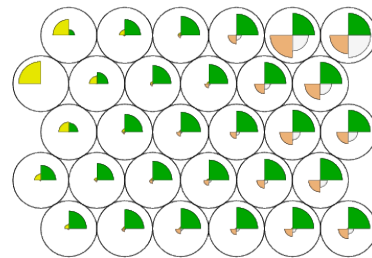
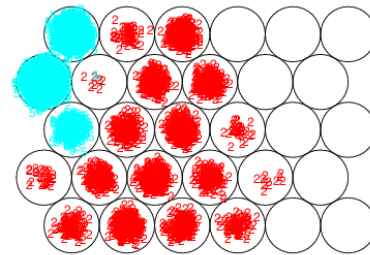


図2.SOMによる全データの図示

図3.SOMによる z_1, z_{18} に属するデータの図示

各ユニットに配属される個体の特徴を示したものを図2に示す。各ユニットの星グラフの4つの軸は反時計回り(右上:初期購買日, 左上:R, 左下:F, 右下:M)に顧客の特徴量に対応する。非優良顧客を表している潜在クラス z_1 (灰色)と優良顧客を表している潜在クラス z_{18} (黒)に属するそれぞれの顧客がユニットにプロットされた結果を図3に示す。図2と図3を対照させることでクラスの特徴を図で可視化することができる。非優良顧客は図3の左上にプロットされ, 初期購買日(First)が全体的に小さく, Rが全体的に大きく, FとMが小さいという特徴をもつ。また, 優良顧客は図3の中央左側にプロットされ, 特徴としては初期購買日が全体的に大きく, Rが小さく, Fが相対的に大きい。このように, 潜在クラスごとの特徴をSOMによって可視化することができる。

6 まとめと今後の課題

本研究では, 顧客の購買履歴データの分析を目的としたRFM特徴量に基づく潜在クラスモデルを拡張し, 初期購買日を考慮した潜在クラスモデルを提案した。また, 自己組織化マップにより, 潜在クラスごとの特徴を可視化する方法を示した。今後の課題として, 潜在クラスごとの購買傾向(購買アイテム)をSOMにより可視化する方法が挙げられる。

参考文献

- [1] 張倩, 三川健太, 後藤正幸, “購買履歴を用いたRFM分析に基づく顧客分析に関する一考察,” 平成27年度日本経営工学会春季大会予稿集, pp. 130–131, 2015.
- [2] T. Kohonen. *Self-Organizing Maps*, Berlin, Springer, vol.30, 1995.