

時系列的な変化を考慮した消費者行動に関する研究～顧客関係管理における活用～

プロフィットエンジニアリング研究

5215F011-9 鈴木慎将
指導教員 大野高裕

A Study of Consumer Behavior Considering Dynamic Changes ～Utilization in Customer Relationship Management～

SUZUKI Shinsuke

1. はじめに

現在、多くの小売店において“優良顧客”が重要となっている。なぜならば、顧客の一部に当たる優良顧客が小売店の売上の多くを創出しているためである。小売店における優良顧客とは、売上や利益、来店頻度等が特に多い顧客のことである。優良顧客識別に用いられる代表的な方法としてRFM分析が挙げられる。RFM分析を用いた優良顧客識別とは、Recency（直近の購買日）、Frequency（観測期間中の購買回数）、Monetary（観測期間中の総購買金額）という3つの値（以下、RFM値）、またはこれらの一部の値に応じて顧客に点数を付け、その点数の順位によって優良顧客を識別する方法である[1]。しかし、マーケティング領域の研究においては、顧客生涯価値¹（以下、CLV）という概念に基づく長期的視野での優良顧客の識別が重要といわれている。しかしながら、RFM値を用いた顧客の順位において、必ずしもCLVが高い顧客が上位になるとは限らない。したがって、優良顧客の識別をCLVの観点から分析する必要がある。加えて、購買行動は消費者間で異質であり、また同一の消費者であっても時系列的に変化する[2]。したがって優良顧客識別のための顧客評価は消費者異質性、時系列的な変化を捉え、逐次的に行う必要がある。そうでなければ、優良顧客向け施策を行なった後の購買行動の変化も捉えられないと考えられる。

本研究ではRFM値を用いて、消費者購買行動の推定、CLVに基づく優良顧客の識別の分析を行なう。このような領域の代表的な研究として阿部[3]が挙げられる。阿部は、CLVに基づく分析のためには、モデルを用いて購買頻度、生存期間、平均購買単価を推定する必要があると述べている。しかし、阿部においては購買行動の時系列的な変化が考慮されていない。また、阿部における計算手法は逐次的な評価に用いるには適していない。そこで本研究では、時系列的な変化を捉え、逐次的な評価を可能とするモデルを構築し、顧客の購買頻度、生存期間、平均購買単価を推定する。その上で、CLVに基づく優良顧客の識別の分析を行なう。

2. 従来研究

2.1. モデル

消費者のRFM値に着目したモデルを構築し、CLVに基づいた分析をした代表的研究として阿部が挙げられる。阿部

はRFM値を用いて消費者の購買頻度、平均購買単価、生存期間を推定するモデルを構築している。これらのパラメータに対する阿部における仮定とモデル式を次に示す。まず、消費者の購買は生存期間中ポアソン・プロセスに従うと仮定し、消費者*i*が観測期間中の最初の購買から時点*T*までにリピート購買する回数が x_i である確率 $P[x_i]$ は

$$P[x_i] = \frac{(\lambda_i T)^{x_i}}{x_i!} e^{-\lambda_i} \quad \text{if } \tau_i \geq T \quad (1)$$
$$P[x_i] = \frac{(\lambda_i \tau_i)^{x_i}}{x_i!} e^{-\lambda_i \tau_i} \quad \text{other}$$

と表される。ただし、 λ_i は購買頻度パラメータ、 τ_i は生存期間である。生存期間は指数分布に従うと仮定し、生存期間が τ_i である確率密度は、次のように表される；

$$P[\tau_i] = \mu_i e^{-\mu_i \tau_i}. \quad (2)$$

ただし、 μ_i はパラメータである。式(2)を用いて、観測期間*T*よりも生存期間 τ_i が長い確率は

$$P[\tau_i > T] = \frac{1}{1 + \frac{\mu_i}{\lambda_i + \mu_i} (e^{(\lambda_i + \mu_i)(T - t_{x,i})} - 1)} \quad (3)$$

と定義することで、確率的に式(1)が定まり、推定における尤度計算に用いることができる。ただし、 $t_{x,i}$ は直近購買時点である。また、購買金額は対数正規分布に従うと仮定し、*n*回目の購買金額 $s_{i,n}$ は次のように表される；

$$P[s_{i,n}] = \frac{1}{\sqrt{2\pi\omega_i^2}} \exp\left(-\frac{(\log(s_{i,n}) - \log(\eta_i))^2}{2\omega_i^2}\right). \quad (4)$$

ただし、 η_i は平均購買単価パラメータ、 ω_i^2 は分散パラメータである。パラメータ λ_i, η_i, μ_i の事前分布は平均ベクトルが $\mathbf{B}d_i$ 、分散共分散行列 $\mathbf{\Gamma}_0$ である多変量の対数正規分布に従うと仮定している。ただし、 \mathbf{B} はパラメータベクトル、 d_i は切片を含む消費者属性ベクトルである。

未知パラメータ $\lambda_i, \mu_i, \eta_i, \mathbf{\Gamma}_0, \mathbf{B}$ はMarkov Chain Monte Carlo method（以下、MCMC）により推定する。購買金額の分散パラメータ ω_i^2 は過去のデータより推定した値を用いて、既知として扱っている。

本研究の目的に対して、阿部のモデルにおいては消費者ごとのパラメータ λ_i, μ_i, η_i が時系列的に一定である。しかし、実際の消費者の購買行動は時系列的に変化すると考えられる[2]。そのため、これらのパラメータにも時系列変化を仮

¹その顧客が店舗滞在期間中にもたらす期待総売上

定し、消費者の購買行動の時系列的な変化を捉える必要がある。したがって、モデル化においてはこれらの値に対して時系列的な変化を仮定し、推定を行なう必要がある。

3. 提案手法

本研究では、消費者ごとに購買頻度、平均購買単価に時系列的な変化を仮定し、逐次的な推定を行なうモデルを構築し、これらの値を用いた CLV の推定、優良顧客の識別や消費者購買行動に関する分析を行なう。

3.1. モデル定式化

モデルの構築および推定は消費者ごとに行なうことを想定する。また、時系列的な変化を購買頻度と平均購買単価に仮定する。そのため、時系列な変化を捉えるために T' 単位で区切られた期間を表わす添え字 t を用いて、これらのパラメータは $\lambda_{i,t}, \eta_{i,t}$ と表わされる。すなわち、購買頻度と平均購買単価は期間 t ごとに変化すると仮定する。

加えて、本研究では、購買金額はガンマ分布に従うと仮定し、 n 回目の購買金額が $s_{i,n}$ である確率は

$$P[s_{i,n}] = s_{i,n}^{k_{i,t}-1} \frac{e^{-s_{i,n}/l_{i,t}}}{\Gamma(k_{i,t})l_{i,t}^{k_{i,t}}} \quad (5)$$

と表す。ただし、 $k_{i,t}, l_{i,t}$ はガンマ分布の形状を表すパラメータであり、 $\eta_{i,t} \equiv k_{i,t}l_{i,t}$ として推定を行なう。分散 $\omega_i^2 \equiv k_{i,t}l_{i,t}^2$ は実際の購買データにおける分散の値を用いる。

期間 t ごとにモデルを構築した場合、購買する回数 $x_{i,t}$ が 0 になる場合が過剰に生じることが想定される。しかしながら、従来のポアソン分布では購買頻度が 0 に寄って小さくなり、0 が突出しているという特徴を捉えることができない。そのため、本研究では Zero-Inflated Poisson[4] を用いて、消費者 i が期間 t に購買する回数が $x_{i,t}$ である確率 $P[x_{i,t}]$ を

$$P[x_{i,t}] = P_{0,i}y_{i,t} + (1 - P_{0,i})P'[x_{i,t}], \quad (6)$$

$$P'[x_{i,t}] = \begin{cases} \frac{(\lambda_{i,t}T_t)^{x_{i,t}}}{x_{i,t}!} e^{-\lambda_{i,t}T_t} & \text{if } \tau_t > T_{i,t}, \\ \frac{(\lambda_{i,t}\tau_{i,t})^{x_{i,t}}}{x_{i,t}!} e^{-\lambda_{i,t}\tau_{i,t}} & \text{otherwise} \end{cases}$$

と表現する。ただし、 $P_{0,i}$ は $x_{i,t} = 0$ である確率を増加させる補正確率、 $y_{i,t}$ は $x_{i,t} = 0$ ならば 1、そうでないならば 0 となるダミー変数である。また、 τ_i は消費者 i の生存期間、 $\tau_{i,t}$ は期間 t 開始時点における残りの生存時間、 T_t は期間 t の日数、 $T_{i,t}$ は期間 t 終了時点における観測期間である。そして、CLV は、次のように計算される；

$$\begin{aligned} \text{CLV} &= \int_0^\infty (\text{時点 } t \text{ での期待売上}) \times (\text{生存関数}) \\ &\quad \times (\text{時間に関する割引率}) dt \\ &= \int_0^\infty \lambda(1 - P_{0,i})\eta \times e^{-\mu} \times e^{-\log(1+d)\frac{t}{365}} dt \\ &= \frac{\lambda(1 - P_{0,i})\eta}{\mu + \log(1+d)/365}. \end{aligned}$$

ただし、 d は時間に対する割引率を表す。

3.2. 推定方法

本研究では Particle Markov Chain Monte Carlo method (以下、PMCMC) により、パラメータの推定を行なう。PMCMC とは、時変パラメータの推定を Particle filter の枠組みで行ない、時変パラメータ以外の未知パラメータ (以下、静的パラメータ) の推定を MCMC により行なう手法である。PMCMC により、時変パラメータの逐次的な推定が可能になる。加えて、パラメータ $\mu_i, P_{0,i}$ のような、モデルにおいて時変ではないパラメータも一般的な Particle filter と比較してより効率よく、正確に推定することができる。

Particle filter とは、非線形なモデルでも推定することのできる、状態推定法の 1 つである。システム内部状態 (以下、状態)、状態の時間変化規則、システムに対する入力情報によりその出力の挙動が定まる動的システムに対し、状態の分布を逐次的に推定する手法が状態推定法である。Particle filter による状態推定ではまず推定したい状態の分布を、粒子と呼ばれる N^2 個の状態の実現値により近似する。次に、状態の時間変化規則に従って、状態の分布を時間更新する。そして観測値を得た際に、観測モデルによる尤度評価に基づいて事後分布を得ることで、状態分布を真の分布に近づける。このように多数の粒子を用いて状態推定する手法が Particle filter である。本研究の購買行動モデルに対しては、動的システムを消費者、システムの状態を時変パラメータ $\lambda_{i,t}, \eta_{i,t}$ 、観測値を消費者の購買行動として適用し、観測モデルとして式 (5)、(6) を用いる。時変パラメータの時間変化規則に関しては、平滑化事前分布を仮定して次のように表わす；

$$\log\left(\frac{\lambda_{i,t}}{1 - \lambda_{i,t}}\right) = \log\left(\frac{\lambda_{i,t-1}}{1 - \lambda_{i,t-1}}\right) + v_{\lambda,i,t}, \quad (7)$$

$$\log(\eta_{i,t}) = \log(\eta_{i,t-1}) + v_{\eta,i,t}. \quad (8)$$

ただし、 $v_{\lambda,i,t}, v_{\eta,i,t}$ は確率の変動項である。式 (7) と式 (8) の変数変換は $0 \leq \lambda_{i,t} \leq 1 \forall (i,t), 0 \leq \eta_{i,t} \forall (i,t)$ という定義域を持つためである。 $v_{\lambda,i,t}, v_{\eta,i,t}$ はそれぞれ独立に、

$$v_{\lambda,i,t} \sim N(0, \tau_{\lambda,i}^2), \quad (9)$$

$$v_{\eta,i,t} \sim N(0, \tau_{\eta,i}^2) \quad (10)$$

に従う。ただし、 $\tau_{\lambda,i}^2, \tau_{\eta,i}^2$ は分散パラメータである。

また、MCMC により静的パラメータ $\tau_{\lambda,i}^2, \tau_{\eta,i}^2, \mu_i, P_{0,i}$ の推定を行なう。MCMC とはモンテカルロ法による推定を、マルコフ連鎖を利用して行なう方法である。本研究では、MCMC におけるサンプリングの手法としてメトロポリス・ヘイスティングス・アルゴリズムを用いる。始めに、MCMC においては繰り返しごとに、前のサンプルから次のサンプルの候補値を得る提案分布 $q(\beta | \beta^{(itt)})$ を定義する。ただし、 β は推定対象のパラメータ ($\tau_{\lambda,i}^2, \tau_{\eta,i}^2, \mu_i, P_{0,i}$)、 itt は実施されたイテレーション回数とする。そして、その提案分布

² N は 10,000,100,000 といった値が用いられる

表 1. 本研究と従来研究の比較

項目	誤差二乗和	阿部	本研究
購買回数	総和	4.02×10^3	3.64×10^3
	平均	6.71	6.06
平均購買金額	総和	1.02×10^9	0.87×10^9
	平均	1.70×10^6	1.45×10^6

$q(\beta | \beta^{(itt)})$ を用いて候補値を取得し、その候補値の採択確率を計算する。採択確率は

$$\alpha = \begin{cases} \min \left[\frac{L(\beta)p(\beta)q(\beta^{(itt)}|\beta)}{L(\beta^{(itt)})p(\beta^{(itt)})q(\beta|\beta^{(itt)})}, 1 \right] & \text{分母} > 0 \\ 1 & \text{分母} = 0 \end{cases} \quad (11)$$

と表す。ただし、 $L(\beta)$ は β における尤度、 $p(\beta)$ は β の事前分布における確率密度である。この採択確率に基づき、得られた候補値 β または $\beta^{(itt)}$ の値が、次のサンプル $\beta^{(itt+1)}$ の値として採択される。このようなサンプリングを決められた回数繰り返す、初期値の影響を受けていると考えられる期間（バーンイン期間）のサンプルを除くことで、推定したい分布をモンテカルロ近似するサンプル群を得る。

4. 検証・考察

4.1. 使用データ

本研究では、GMS³の ID 付き POS データにおける食品の購買データを用いて検証を行なう。ランダムにサンプリングされた 1 店舗を対象店舗とする。そして、対象店舗における 200 世帯をサンプルとする。データの期間は 2 年間で、最初の 3 カ月を各変数の初期値の算出・推定に、次の 1 年半をパラメータの推定に、残りの 3 カ月を予測・検証に使用する。

4.2. 検証

本研究の推定における設定は

1. Particle filter における粒子数は 10,000 個
2. MCMC におけるイテレーション回数は 30,000 回
3. MCMC におけるバーンイン期間は初期 10,000 回
4. T' は 1 カ月

となっている。MCMC における収束判定は Geweke[5] のテストによって行なう。本研究では、従来研究との比較によって数理モデルの検証を行なう。比較には、推定期間 3 カ月における、各月の購買回数・平均購買金額の予測値と実測値の二乗誤差を用いる。その結果を表 1 に示す。表 1 より、どちらの予測精度も本研究の方が優れている。そのため、購買行動に時系列的な変化を仮定し、逐次的な評価を可能にした本研究の提案手法の妥当性が証明されたと考える。

4.3. 考察

推定された CLV を用いて、RFM 分析を用いた優良顧客識別に関する分析を行なう。具体的には、横軸に消費者の

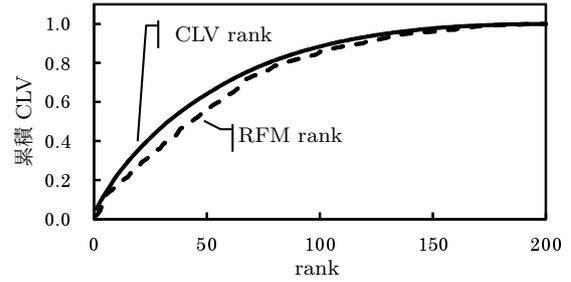


図 1. RFM 分析による順位に基づく優良顧客識別の分析順位、縦軸にその順位までの消費者の累積 CLV をプロットしたゲインチャートを作成する。横軸の順位に「推定された CLV の高さの順位」と「RFM 各値のランキングの平均値が高さの順位」を用いた 2 つの曲線をプロットしたものが図 1 である。この図において、「推定された CLV の高さの順位」を用いた曲線（以下、CLV 曲線）は得られると期待される累積 CLV の最大値を示す。そのため、CLV 曲線に近いほど望ましい。図 1 より、「RFM 各値のランキングの平均値が高さの順位」の曲線は、CLV 曲線とおおよそ近い動きをしている。したがって、優良顧客識別に関して、ある程度推定された CLV が高い消費者を識別できたと考えられる。

次に、パラメータ $\tau_{\lambda,i}^2, \tau_{\eta,i}^2$ に着目し、各消費者のパラメータの時系列的な変化の程度を考察する。パラメータ $\tau_{\lambda,i}^2, \tau_{\eta,i}^2$ は PMCMC において、購買頻度と平均購買単価を変化させる確率の変動項の大きさに影響する。そのため、各消費者におけるこれらの時系列的な変化の度合いを端的に表すと考えられる。購買頻度と平均購買単価に関して、その値と対応するパラメータの値を散布図にプロットしたものが図 2, 3 となる。図 2 では時系列的な変化の度合いが 0、すなわち購買頻

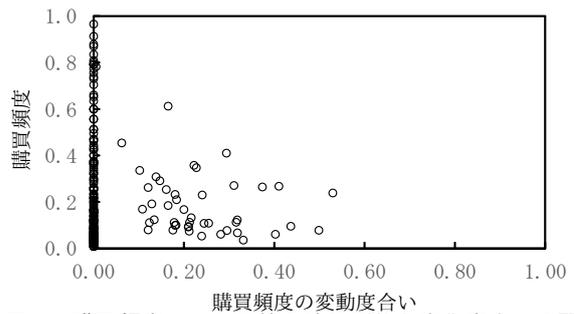


図 2. 購買頻度における値と時系列的な変化度合いの関係

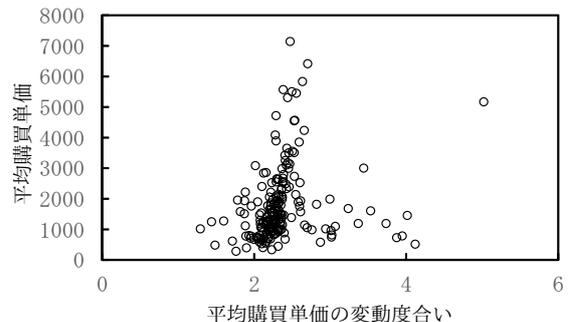


図 3. 平均購買単価における値と時系列的な変化度合いの関係

³general merchandise store

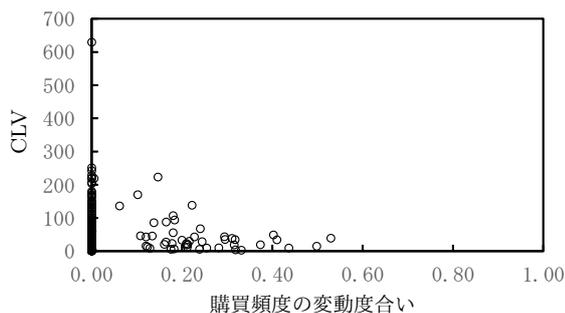


図 4. 購買頻度における時系列的な変化度合いと CLV の関係

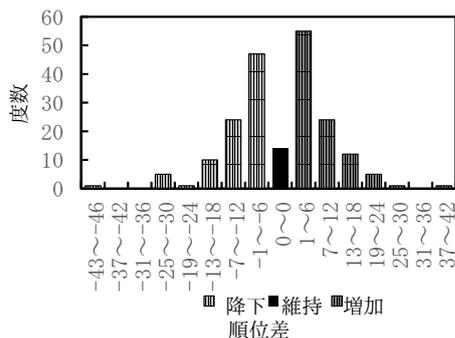


図 5. 顧客ランクの順位差ごとの消費者度数分布

度が時系列的に変化しない消費者が多くいることが分かる。これらの消費者は、非常に安定した頻度で来店し、購買していることが分かる。一方、図 3 においては変化の度合いが 0 となる消費者は存在しない。ただし、一部相対的な変動度合いが大きく、かつ平均購買単価の値も大きな消費者が存在する。この消費者の内、推定された平均購買単価の高い消費者に対しては、平均購買単価の変化の要因を分析し（季節性、等）、平均購買単価を高い水準で維持できれば、多くの売上を得られる可能性がある。ここで、購買頻度における時系列的な変化がない消費者に着目し、購買頻度における変化度合いと CLV の関係を示した散布図が図 4 となる。購買頻度における時系列的な変化がない消費者において、その一部は相対的な CLV が高い。したがって、これらの消費者は「購買頻度が安定的で、かつ CLV の高い重要な消費者」であることがわかる。そのため、これらの消費者に対しては、

- 購買頻度は安定的であるため、購買頻度を増やすための特別な施策は基本的に行なわない
- 新規競合出店等、環境の変化に連動して離脱しないよう注意をする

といったアプローチを行なうことが有効であると考えられる。

この内、順位が降下した消費者に着目し、購買データから購買の変化を分析する。順位が降下した消費者において、商品分類ごとの総購買金額の平均値を集計したものが図 6 である。ただし、平均総購買金額の上位 10 商品分類における値を表示する。推定期間におけるパラメータを用いて計算された CLV をもとに順位を計算する。その推定期間における最

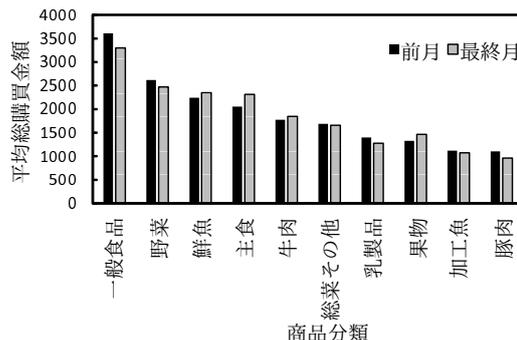


図 6. 順位低下消費者の大分類ごと総購買金額の平均値

終月と最終月の前月における順位差（最終月の前月の順位 - 最終月の順位）を端的に表したものが図 5 である。図 6 において、特に減少額は“一般食品”と“野菜”が大きいことがわかり、その減少率はそれぞれ -8.5%，-5.7% である。この“一般食品”は調味料、飲料、レトルト食品といった商品が含まれるカテゴリである。これらの商品に関して、総購買金額が減少している状態が一時的なものでないならば、販売促進を行なう必要があると考えられる。

5. おわりに

本研究では、時系列的な変化を捉え、各消費者の購買頻度、生存期間、平均購買単価を逐次的に推定できる購買行動モデルを構築した。また、それらの値を用いて、CLV の値を推定することで、優良顧客の識別に関して CLV に基づく分析ができた。今後の課題としては、対象とする店舗・消費者の拡大、購買行動の時系列的な変化に対する季節性の考慮が挙げられる。

6. 謝辞

本研究は NEDO 委託事業「人間と相互理解できる次世代人工知能技術の研究開発」の支援を受けている。

参考文献

- [1] フィリップ・コトラー：「コトラーのマーケティング・マネジメント」月谷真紀訳、株式会社ピアソン・エデュケーション、pp.801-802 (2001)
- [2] 佐藤忠彦、樋口知之：ビッグデータ時代のマーケティング、pp.13-19、講談社 (2013)
- [3] 阿部誠：”RFM 指標と顧客生涯価値：階層ベイズモデルを使った非契約型顧客関係管理における消費者行動の分析”，日本統計学会誌，Vol.41，pp.51-81 (2011)
- [4] Lambert, D: “Zero-inflated poisson regression, with an application to defects in manufacturing”, *Technometrics*, Vol.34, No.1, pp.1-14 (1992)
- [5] Geweke, J.: “Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments”, *Bayesian Statistics*, Vol.4, pp.169-193 (1992)