2Q1-J-2-02

RFM 指標と顧客特性に基づく LTV 予測モデル

LTV Prediction Based on RFM and Customer Characteristics

今井 優作　田嶋 優樹
Yusaku Imai　Yuki Tajima

株式会社電通デジタル
Dentsu Digital Inc.

Lifetime Value (LTV) as we know it, is an important indicator of customer evaluation. To build long-term relationships with the right customer, it is important to predict LTV with increasingly higher accuracy levels. Once we attain that, we would be able to communicate with them through appropriate marketing actions. While predicting LTV in a non-contractual setting, three indicators, namely; Recency, Frequency and Monetary Value (RFM) are widely used. RFM is used as an indicator of customers' buying behaviour on the whole, however normally dimensions like demographics are not considered. In this paper, we propose a model for predicting LTV based not just on RFM, but also other customer characteristics. To support our proposal and its effectiveness we have also provided the details of the experiments, their outputs and our inference using a real dataset.

1. はじめに

マーケティングにおける顧客評価のための重要な指標として顧客生涯価値 (LTV; Lifetime Value) がある．非契約型サービスにおいて，LTV は顧客の生存期間, 購買頻度, 1 回あたりの購買金額の掛け合わせで表される [Schmittlein 94]．LTV を高精度に予測できるようになると，適切なマーケティング活動を通して優良顧客との長期的な関係構築が可能となる．

LTV 予測において，Recency(直近の購買からの経過日数), Frequency(観測期間中の購買回数), Monetary(平均購買金額) の 3 指標が広く使われている．RFM 指標は顧客の購買特性を集約した指標として有用であるが，性別や年代といった顧客自体の特性は考慮していない．本稿では，RFM 指標に加えて顧客特性も加味した予測モデルを提案し，RFM 指標のみを用いた従来法よりも高い精度で LTV 予測が可能なことを示す．

2. LTV 予測

RFM 指標を用いた LTV 予測モデルの代表的な手法として BTYD(Buy Till You Die) モデルがある [Schmittlein 87]．観測終了時刻を T とすると，$(T, T+t]$ での予測 LTV は，T での顧客生存確率 $\times$ $(T, T+t]$ での期待購買回数 $\times$ 1 回あたりの期待購買金額となり，第一項, 第二項の購買回数は BG/NBD モデル [Fader 05b]，第三項の購買金額は GG モデル [Fader 05a] を用いて求められる．

BG/NBD モデルでは，購買回数にパラメータ λ のポアソン分布，顧客の離脱確率にパラメータ ρ の幾何分布を仮定する．また，λ はパラメータ r, β のガンマ分布，ρ はパラメータ a, b のベータ分布を事前分布とする．$(0, T]$ での購買回数が x，最終購買時刻が t_x のとき，T での顧客生存確率は

$$P(\text{active}|x, t_x, T; r, \beta, a, b) = \frac{1}{1 + \delta_{x>0} \frac{a}{b+x-1} \left(\frac{\beta+T}{\beta+t_x}\right)^{r+x}}, \tag{1}$$

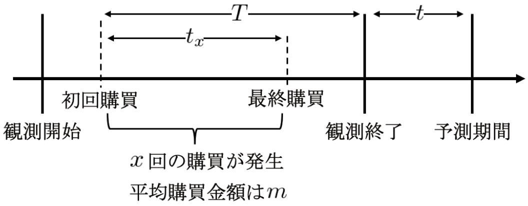

図 1: RFM 指標の概要図

$(T, T+t]$ での期待購買回数は

$$\begin{aligned} &E(X(t)|x, t_x, T; r, \beta, a, b) \\ &= \frac{a+b+x-1}{a-1}\left[1 - \left(\frac{\beta+T}{\beta+T+t}\right)^{r+x}\right. \\ &\quad \left. \cdot\ {}_2F_1\left(r+x, b+x; a+b+x-1; \frac{t}{\beta+T+t}\right)\right] \end{aligned} \tag{2}$$

となる．ここで，δ はデルタ関数, ${}_2F_1$ はガウスの超幾何関数を表す．

GG モデルでは，購買金額にパラメータ τ, ν のガンマ分布を仮定し，ν はパラメータ q, ξ のガンマ分布を事前分布とする．$(0, T]$ での購買回数が x，平均購買金額が m のとき，期待購買金額は

$$E(M|m; \tau, q, \xi) = \frac{\tau(\xi + mx)}{\tau x + q - 1} \tag{3}$$

となる．

式 (1)(2)(3) の導出過程については巻末付録に示す．

3. 提案法

本稿では，2. で示した BG/NBD モデルと GG モデルを基に，RFM 指標に加えて顧客特性も加味した LTV 予測モデルを提案する．

顧客特性を $\boldsymbol{d} = \{d_k\}$ とする．BG/NBD モデルのパラメータ r, β, a, b および GG モデルのパラメータ τ, q, ξ はいずれも

連絡先: 今井優作, 株式会社電通デジタル, imai.yu@dentsudigital.co.jp

> 0 となるため，モデルパラメータは $\vartheta = \exp(\langle \boldsymbol{w}_\vartheta, \boldsymbol{d} \rangle), \vartheta \in \{r, \beta, a, b, \tau, q, \xi\}$ とする．ここで，$\langle \cdot, \cdot \rangle$ は 2 つのベクトルの内積を表す．

従来法ではパラメータ $r, \beta, a, b, \tau, q, \xi$ を推定するが，提案法では以下の最適化問題を解いて重みパラメータ $\boldsymbol{w}_\vartheta$ を求める．

$$\arg\min_{\boldsymbol{w}_\vartheta} L + \frac{\mu}{2} \left(\sum_\vartheta \|\boldsymbol{w}_\vartheta\|_2^2 \right) \tag{4}$$

ここで，μ は正則化パラメータ, $\|\boldsymbol{w}\|_2^2$ は L2 ノルム, L は negative log-likelihood を表す．

$\boldsymbol{w}_\vartheta$ の勾配は連鎖律により

$$\frac{\partial L}{\partial w_{\vartheta,k}} = \frac{\partial L}{\partial \vartheta} \cdot \frac{\partial \vartheta}{\partial w_{\vartheta,k}} = \frac{\partial L}{\partial \vartheta} \cdot \vartheta d_k \tag{5}$$

となる．ここで，BG/NBD モデルの尤度

$$\begin{aligned} \ell_{\mathrm{BG/NBD}} = &\frac{B(a, b+x)}{B(a,b)} \frac{\Gamma(r+x)\beta^r}{\Gamma(r)(\beta+T)^{r+x}} \\ &+ \delta_{x>0} \frac{B(a+1, b+x-1)}{B(a,b)} \frac{\Gamma(r+x)\beta^r}{\Gamma(r)(\beta+t_x)^{r+x}} \end{aligned} \tag{6}$$

より

$$\begin{aligned} \frac{\partial L_{\mathrm{BG/NBD}}}{\partial r} = &\Psi(r) - \Psi(r+x) - \log\left(\frac{\beta}{\beta+T}\right) \\ &- \frac{\delta_{x>0}\left(\frac{a}{b+x-1}\right)\left(\frac{\beta+T}{\beta+t_x}\right)^{r+x}\log\left(\frac{\beta+T}{\beta+t_x}\right)}{1+\delta_{x>0}\left(\frac{a}{b+x-1}\right)\left(\frac{\beta+T}{\beta+t_x}\right)^{r+x}}, \end{aligned} \tag{7}$$

$$\begin{aligned} \frac{\partial L_{\mathrm{BG/NBD}}}{\partial \beta} = &-\frac{r}{\beta} + \frac{r+x}{\beta+T} \\ &- \frac{\delta_{x>0}\left(\frac{a}{b+x-1}\right)\left(\frac{\beta+T}{\beta+t_x}\right)^{r+x}\left(\frac{r+x}{\beta+T}\right)\left(1-\frac{\beta+T}{\beta+t_x}\right)}{1+\delta_{x>0}\left(\frac{a}{b+x-1}\right)\left(\frac{\beta+T}{\beta+t_x}\right)^{r+x}}, \end{aligned} \tag{8}$$

$$\begin{aligned} \frac{\partial L_{\mathrm{BG/NBD}}}{\partial a} = &-\Psi(a+b) + \Psi(a+b+x) \\ &- \delta_{x>0}\frac{\left(\frac{\beta+T}{\beta+t_x}\right)^{r+x}}{\left(1+\left(\frac{a}{b+x-1}\right)\left(\frac{\beta+T}{\beta+t_x}\right)^{r+x}\right)(b+x-1)}, \end{aligned} \tag{9}$$

$$\begin{aligned} \frac{\partial L_{\mathrm{BG/NBD}}}{\partial b} = &-\Psi(a+b) + \Psi(a+b+x) + \Psi(b) - \Psi(b+x) \\ &+ \delta_{x>0}\frac{\left(\frac{\beta+T}{\beta+t_x}\right)^{r+x}\left(\frac{a}{b+x-1}\right)}{\left(1+\left(\frac{a}{b+x-1}\right)\left(\frac{\beta+T}{\beta+t_x}\right)^{r+x}\right)(b+x-1)}, \end{aligned} \tag{10}$$

GG モデルの尤度

$$\ell_{\mathrm{GG}} = \frac{\Gamma(\tau x+q)}{\Gamma(\tau x)\Gamma(q)} \frac{m^{\tau x-1}x^{\tau x}\xi^q}{(\xi+mx)^{\tau x+q}} \tag{11}$$

より

$$\frac{\partial L_{\mathrm{GG}}}{\partial \tau} = \left[-\Psi(\tau x+q) + \Psi(\tau x) - \log\left(\frac{mx}{\xi+mx}\right)\right] x, \tag{12}$$

$$\frac{\partial L_{\mathrm{GG}}}{\partial q} = -\Psi(\tau x+q) + \Psi(q) - \log\left(\frac{\xi}{\xi+mx}\right), \tag{13}$$

$$\frac{\partial L_{\mathrm{GG}}}{\partial \xi} = -\frac{q}{\xi} + \frac{\tau x+q}{\xi+mx} \tag{14}$$

となる．ここで，B はベータ関数, Γ はガンマ関数, Ψ はディガンマ関数を表す．

表 1: MAE の比較

	従来法	提案法
期待購買回数	2.852	2.722
期待購買金額 ($\times 10^3$)	4.082	4.046
予測 LTV ($\times 10^3$)	15.180	14.457

表 2: RMSLE の比較

	従来法	提案法
期待購買回数	0.884	0.845
期待購買金額	1.045	1.035
予測 LTV	6.134	6.023

4. 実験

提案法の有効性を評価するための実験を行った．実験には，セブン＆アイ・ホールディングスの通販サイト omni7[*1] における，2016 年 10 月 1 日から 2018 年 9 月 30 日までの購買データを用いた．購買データには，4,930,238 の顧客 ID と 22,500,570 の購買情報が含まれる．また，提案法における顧客特性として，性別, 年代, 地域の 3 変数を用いた．

2017 年 9 月 30 日までの RFM 指標を用いてモデルの学習を行い，2017 年 10 月 1 日から 1 年後までの期待購買回数 (式 (1)×(2)), 期待購買金額 (式 (3)), および予測 LTV(式 (1)×(2)×(3)) を求めた．評価尺度には平均絶対誤差 (MAE) と対数平方平均二乗誤差 (RMSLE) を用いた．実験結果を表 1, 2 に示す．表より，いずれの評価指標においても提案法の予測精度が向上していることがわかる．また，予測 LTV の改善幅が期待購買回数, 期待購買金額よりも大きいことから，顧客特性を加味することで BG/NBD モデルと GG モデルの予測精度が同時的に向上したと考えられる．

5. おわりに

本稿では，非契約型サービスにおける LTV 予測モデルを提案した．提案法では，顧客の購買特性を示す RFM 指標に加えて顧客特性を加味することで，RFM 指標のみを用いた従来法よりも高い精度で LTV 予測が可能であることを確認した．

先行研究 [阿部 11] において，顧客の離脱確率/購買回数に Pareto/NBD モデル [Schmittlein 87]，購買金額に対数正規分布を仮定し，それらのパラメータの事前分布を多変量対数正規分布とすることで，顧客生存確率, 購買回数, および購買金額の相関構造をモデル化している．そこで，提案法に対しても相関構造を考慮する機構を加え，提案法の有効性に対するさらなる検証，および改善を行う．

*1 `https://www.omni7.jp`

A　BG/NBD モデル

$(0, t]$ での購買回数を x とすると，x は負の二項分布に従う．

$$f(X(t)=x;r,\beta)=\int f(X(t)=x;\lambda)\cdot f(\lambda;r,\beta)d\lambda$$
$$=\frac{\Gamma(r+x)}{\Gamma(r)x!}\left(\frac{\beta}{\beta+t}\right)^r\left(\frac{t}{\beta+t}\right)^x \tag{15}$$

with

$$f(X(t)=x;\lambda)=\frac{(\lambda t)^x\exp(-\lambda t)}{x!}, \tag{16}$$
$$f(\lambda;r,\beta)=\frac{\beta^r\lambda^{r-1}\exp(-\lambda\beta)}{\Gamma(r)} \tag{17}$$

顧客の離脱確率を ρ とすると，j 回目の購買後に離脱する確率はベータ-幾何分布に従う．

$$f(\text{inactive};a,b)=\int f(\text{inactive};\rho)\cdot f(\rho;a,b)d\rho$$
$$=\frac{B(a+1,b+j-1)}{B(a,b)} \tag{18}$$

with

$$f(\text{inactive};\rho)=\rho(1-\rho)^{j-1}, \tag{19}$$
$$f(\rho;a,b)=\frac{\rho^{a-1}(1-\rho)^{b-1}}{B(a,b)} \tag{20}$$

$(0, t]$ での購買回数が x の確率は

$$P(X(t)=x;r,\beta,a,b)$$
$$=\int P(X(t)=x;\lambda,\rho)\cdot f(\lambda;r,\beta)\cdot f(\rho;a,b)d\lambda d\rho$$
$$=\frac{B(a,b+x)}{B(a,b)}\frac{\Gamma(r+x)}{\Gamma(r)x!}\left(\frac{\beta}{\beta+t}\right)^r\left(\frac{t}{\beta+t}\right)^x$$
$$+\ \delta_{x>0}\frac{B(a+1,b+x-1)}{B(a,b)}$$
$$\cdot\left[1-\left(\frac{\beta}{\beta+t}\right)^r\left\{\sum_{j=0}^{x-1}\frac{\Gamma(r+j)}{\Gamma(r)j!}\left(\frac{t}{\beta+t}\right)^j\right\}\right] \tag{21}$$

with

$$P(X(t)=x;\lambda,\rho)$$
$$=P(\text{active after } x\text{th purchase})\cdot P(t_x\le t<t_{x+1})$$
$$+\ \delta_{x>0}P(\text{inactive after } x\text{th purchase})\cdot P(t_x\le t)$$
$$=(1-\rho)^x\frac{(\lambda t)^x\exp(-\lambda t)}{x!}$$
$$+\ \delta_{x>0}\rho(1-\rho)^{x-1}\left[1-\exp(-\lambda t)\sum_{j=0}^{x-1}\frac{(\lambda t)^j}{j!}\right] \tag{22}$$

であり，期待購買回数は

$$E(X(t);r,\beta,a,b)$$
$$=\int E(X(t);\lambda,\rho)\cdot f(\lambda;r,\beta)\cdot f(\rho;a,b)d\lambda d\rho$$
$$=\frac{a+b-1}{a-1}\left[1-\left(\frac{\beta}{\beta+t}\right)^r{}_2F_1\left(r,b;a+b-1;\frac{t}{\beta+t}\right)\right] \tag{23}$$

with

$$E(X(t);\lambda,\rho)=\frac{1}{\rho}(1-\exp(-\lambda\rho t)) \tag{24}$$

となる．

$(0, T]$ での購買回数が x，最終購買時刻が t_x のとき，$(T, T+t]$ での期待購買回数は

$$E(\gamma(t)|x,t_x,T;r,\beta,a,b)$$
$$=P(\text{active}|x,t_x,T;r,\beta,a,b)\cdot E(X(t)|x,t_x,T;r,\beta,a,b) \tag{25}$$

となる．ここで，第一項は

$$P(\text{active}|x,t_x,T;r,\beta,a,b)$$
$$=\int P(\text{active}|x,t_x,T;\lambda,\rho)\cdot f(\lambda,\rho|x,t_x,T;r,\beta,a,b)d\lambda d\rho$$
$$=\frac{1}{1+\delta_{x>0}\frac{a}{b+x-1}\left(\frac{\beta+T}{\beta+t_x}\right)^{r+x}} \tag{26}$$

with

$$P(\text{active}|x,t_x,T;\lambda,\rho)=\frac{(1-\rho)\exp(-\lambda(T-t_x))}{\rho+(1-\rho)\exp(-\lambda(T-t_x))}, \tag{27}$$

$$f(\lambda,\rho|x,t_x,T)$$
$$=(1-\rho)^x\lambda^x\exp(-\lambda T)+\delta_{x>0}\ \rho(1-\rho)^{x-1}\lambda^x\exp(-\lambda t_x), \tag{28}$$

$$f(\lambda,\rho|x,t_x,T;r,\beta,a,b)$$
$$=\frac{f(\lambda,\rho|x,t_x,T)\cdot f(\lambda;r,\beta)\cdot f(\rho;a,b)}{\int f(\lambda,\rho|x,t_x,T)\cdot f(\lambda;r,\beta)\cdot f(\rho;a,b)d\lambda d\rho} \tag{29}$$

第二項は

$$E(X(t)|x,t_x,T;r,\beta,a,b)$$
$$=\int E(X(t);\lambda,\rho)\cdot f(\lambda,\rho|X=x,t_x,T;r,\beta,a,b)d\lambda d\rho$$
$$=\frac{a+b+x-1}{a-1}\left[1-\left(\frac{\beta+T}{\beta+T+t}\right)^{r+x}\right.$$
$$\left.\cdot\ {}_2F_1\left(r+x,b+x;a+b+x-1;\frac{t}{\beta+T+t}\right)\right] \tag{30}$$

である．

B　GG モデル

購買金額 m はガンマ-ガンマ分布に従う．

$$f(M(x)=m;\tau,q,\xi)=\int f(M(x)=m;\tau,\nu)\cdot f(\nu;q,\xi)d\nu$$
$$=\frac{\Gamma(\tau x+q)}{\Gamma(\tau x)\Gamma(q)}\frac{m^{\tau x-1}x^{\tau x}\xi^q}{(\xi+mx)^{\tau x+q}} \tag{31}$$

with

$$f(M(x)=m;\tau,\nu)=\frac{(\nu x)^{\tau x}m^{\tau x-1}\exp(-\nu xm)}{\Gamma(\tau x)}, \tag{32}$$
$$f(\nu;q,\xi)=\frac{\xi^q\nu^{q-1}\exp(-\nu\xi)}{\Gamma(q)} \tag{33}$$

期待購買金額ζはガンマ分布より$\zeta = \frac{\tau}{\nu}$であり，$\zeta = h(\nu;\tau)$とすると

$$f(\zeta;\tau,q,\xi) = \left|\frac{d}{d\zeta}h^{-1}(\zeta;\tau)\right| f(h^{-1}(\zeta;\tau);q,\xi) = \frac{(\tau\xi)^q\zeta^{-q-1}\exp(-\frac{\tau\xi}{\zeta})}{\Gamma(q)} \quad (34)$$

のように逆ガンマ分布となるため，期待値は

$$E(M;\tau,q,\xi) = \frac{\tau\xi}{q-1} \quad (35)$$

となる．

$(0,T]$での購買回数がx，平均購買金額がmのとき

$$f(\nu|m;\tau,q,\xi) = \frac{f(m;\tau,\nu)\cdot f(\nu;q,\xi)}{\int f(m;\tau,\nu)\cdot f(\nu;q,\xi)d\nu} \quad (36)$$

より

$$f(\zeta|m;\tau,q,\xi) = \left|\frac{d}{d\zeta}h^{-1}(\zeta;\tau)\right| f(h^{-1}(\zeta;\tau)|m;\tau,q,\xi) = \frac{(\tau(\xi+mx))^{\tau x+q}\zeta^{-\tau x-q-1}\exp(-\frac{\tau(\xi+mx)}{\zeta})}{\Gamma(\tau x+q)} \quad (37)$$

となるため, 期待購買金額は

$$E(M|m;\tau,q,\xi) = \frac{\tau(\xi+mx)}{\tau x+q-1} \quad (38)$$

である．

参考文献

[Fader 05a] Fader, P. S., Hardie, B. G., and Lee, K. L.: RFM and CLV: Using Iso-Value Curves for Customer Base Analysis, *Journal of Marketing Research*, Vol. 42, No. 4, pp. 415–430 (2005)

[Fader 05b] Fader, P., G. S. Hardie, B., and Lok Lee, K.: “Counting Your Customers” the Easy Way: An Alternative to the Pareto/NBD Model, *Marketing Science*, Vol. 24, pp. 275–284 (2005)

[Schmittlein 87] Schmittlein, D. C., Morrison, D. G., and Colombo, R.: Counting Your Customers: Who Are They and What Will They Do Next?, *Manage. Sci.*, Vol. 33, No. 1, pp. 1–24 (1987)

[Schmittlein 94] Schmittlein, D. and Peterson, R.: Customer Base Analysis: An Industrial Purchase Process Application, *Marketing Science*, Vol. 13, pp. 41–67 (1994)

[阿部 11] 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. シリーズ J, Vol. 41, No. 1, pp. 51–81 (2011)