

# 非負値行列因子分解を用いたプラットフォームビジネスにおける顧客生涯価値予測

蓮本 恭輔<sup>1,a)</sup> 雲居 玄道<sup>1,b)</sup> 後藤 正幸<sup>2,c)</sup>

受付日 2018年10月3日, 採録日 2019年3月5日

**概要:** 近年のオンラインビジネスにおいて、楽天市場やメルカリなどのマーケットプレイス型ビジネスや Apple Pay のような決済サービスなど、消費者と事業者の双方にサービスを提供するビジネスの躍進が目覚ましい。これらはプラットフォームビジネスと呼ばれ、様々な参加者をひきつけ、ビジネスを拡大し、市場を席巻している。プラットフォームビジネスの多くは顧客の新規獲得や利用頻度向上のために、様々なマーケティング施策を行っているが、適切な投資対効果を得るためには自社の顧客が将来どれだけの収益をもたらすのか、優良な顧客がだれなのかを把握することが非常に重要となる。そのため、多くの企業では過去の購買行動に基づき、顧客をセグメント化し、顧客生涯価値を算出している。しかし、それらは過去の購買行動の結果であって、必ずしも将来の購買を約束するものではないため、正しい顧客生涯価値を把握するためには適切な予測モデルが必要になる。本研究では非負値行列因子分解 (NMF) を用いて、顧客の購買パターンを抽出し、顧客生涯価値の予測に適用するモデルを提案する。従来の NMF を予測問題にそのまま適用すると、学習データと予測データはそれぞれ異なるクラスタを生成してしまい、学習に基づく予測ができない。そこで学習結果を予測時の NMF に利用できるようなアルゴリズムの改良方法を示す。また提案手法を実データ分析に適用し、その有効性を検証する。

**キーワード:** 顧客生涯価値 (CLV), プラットフォームビジネス, 非負値行列因子分解 (NMF), RFM 分析

## A Prediction of Customer Lifetime Value in a Platform Business using Nonnegative Matrix Factorization

KYOSUKE HASUMOTO<sup>1,a)</sup> GENDO KUMOI<sup>1,b)</sup> MASAYUKI GOTO<sup>2,c)</sup>

Received: October 3, 2018, Accepted: March 5, 2019

**Abstract:** In the recently online business industry, marketplaces such as Rakuten Ichiba and Mercari in Japan, or payment platforms such as Apple Pay are servicing both consumers and merchants and growing tremendous speed. These businesses, called “platform businesses”, are attracting various participants on their platform and expanding and conquering the markets. Many of the platform businesses are investing various marketing initiatives to acquire new customers and improve customers’ engagements. It is crucial for them to identify the most valuable customers and how much they can contribute to the business so that proper ROI is maintained. As such, many of them are segmenting customers and calculating Customer Lifetime Value basing on the past purchase behaviors. But the CLV from the past purchases are not always the good prediction of future purchases. Therefore, comprehending and calculating Customer Lifetime Value require a proper prediction model. In this research, we propose a model to apply NMF to extract purchase behaviors of customers and use outputs from NMF to predict future Customer Lifetime Value. When conventional NMF is applied to a prediction task, a training data set and a test data set generate different sets of clusters resulting the training result not being reflected to the test result. We showed an adjusted model of NMF to overcome such problem. Finally, we validated our model using actual business data from a platform business.

**Keywords:** customer lifetime value (CLV), platform business, nonnegative matrix factorization (NMF), RFM analysis

## 1. はじめに

近年、オンラインビジネスにおいて最も成長が目覚ましい領域の1つがプラットフォームビジネス [1] である。アマゾンや楽天のイーコマースにおけるプラットフォーム、Apple Pay や Google Pay のような決済プラットフォーム、Airbnb や Uber のような消費者同士を結び付けるプラットフォームなど様々な形がある。これらのビジネスは多くの場合、消費者にサービスを利用してもらうことで収益が入るため、プラットフォームの利用者・利用回数を増やすために様々なマーケティング施策を行っている。同時に適切な投資対効果を得るために顧客をセグメント化し、適切なターゲットを見極め、予算を配分している。そして多くの場合、優良な顧客セグメントの判別、すなわちターゲットの選定は顧客生涯価値 (CLV: Customer Lifetime Value) [2] に基づいて行われる。そのため、CLV の把握は基本戦略やマーケティング施策策定のうえで重要な意味を持つ。一般的に CLV の算出は過去の購買履歴に基づいて行われるが、それは過去の顧客行動の結果であり、必ずしも未来の購入を約束するものではない。ゆえに CLV を正しく把握するためにはなんらかの予測モデルが必要となる。CLV の研究においては、顧客の将来の購入を予測する Pareto/NBD モデル [3] が広く知られており、それをベースにした研究 [4], [5] も行われている。これらのモデルは RFM 指標と呼ばれる直近の購買日や購買回数を入力データとして、予測を行う。また RFM 指標をランダムフォレストのような比較的精度の高い機械学習モデルの説明変数として取り込み、予測を行うこともできる。しかし、これらは簡便な手法である一方で RFM 指標のみではパラメータ数が十分とはいえず、またプラットフォームビジネスに適用した場合、CLV に影響を与える多様な顧客の購買行動や背後にあるプラットフォーム事業者と導入店舗の手数料関係などは考慮できない。顧客がどのような店舗からどのような頻度で購入しているといった顧客の購買パターンが CLV に大きな影響を与えることから、これらの購買履歴データを予測器の入力変数として活用することを考える。しかし、取扱店舗数は非常に多く、かつスパースなデータであることから、これらをすべて入力変数としてそのまま用いても高精度な予測モデルを構築できないという問題が生じる。そこで本研究では、非負値行列因子分解 (NMF: Non-negative Matrix

Factorization) [6], [7] を用いて次元を縮約し、顧客の購買パターンを抽出した上で RFM 指標と組み合わせて、ランダムフォレスト回帰による CLV 予測に適用する。

高次元のデータを NMF で次元圧縮し、異なる機械学習のモデルと組み合わせる方法は様々な形で用いられている [8], [9]。しかし NMF の行列分解の結果は学習データに対するものであるため、学習データにはなかった新規入力データに対する予測に適用する場合は、どのように次元縮約した特徴量ベクトルを作成するかを再考しなければならない。すでに学習フェーズにおいて、NMF で次元縮約した特徴ベクトルを用いて、機械学習による予測器を構築しているため、新たに新規入力データも加えて NMF を実行すると、異なる特徴空間が構成されてしまい、予測器の入力として使うことができなくなってしまう。すなわち、新規データの次元縮約の結果が NMF を学習したときと同じ特徴空間でマッピングされるようにする必要がある。そこで本研究では学習データに対して NMF で特徴抽出した際に得られた部分行列を活用し、新規データを同じ特徴空間にマッピングする方法を構成し、この問題を解決する。最後に実データ分析への適用についても説明したい。

## 2. 準備

### 2.1 関連研究

CLV の予測や機械学習を用いた購買分析は様々な形で研究がなされているが、CLV のモデル化や予測手法の観点から関連する研究を以下にまとめる。

#### 2.1.1 顧客生涯価値と RFM 分析

顧客生涯価値 (CLV) は個々の顧客が対象の事業者に生涯にわたってもたらす収益を指す。近年、顧客関係管理 (CRM: Customer Relationship Management) の概念が広がり、顧客 1 人 1 人との関係が重視されるようになって、CLV はその中心的な概念の 1 つとして用いられている [2]。企業経営において CLV を把握することは、マーケティング施策決定のみならず、経営計画や事業戦略の策定において重要な意味を持つ。これは CLV の大きさが将来の企業の収益を左右するためである。一般的に顧客ごとの CLV は将来の利益を現在価値に割り引くことで算出される。CLV の計算期間を  $T$ 、第  $p$  期の利益を  $m_p$ 、1 期ごとの割引率を  $i$  とすると、CLV は式 (1) で計算される。

$$CLV = \sum_{p=0}^T \frac{m_p}{(1+i)^p} \quad (1)$$

しかし実際には未来の購入からもたらされる利益は不明であるため、CLV を把握するにはなんらかの予測モデルが必要となる。CLV のモデル化や予測は様々な形で試みられており [4], [5], [10], [11], [12]、企業による実データを使った実験 [13], [14] も行われている。また CLV の予測では RFM 分析で使用される最新購買日 (Recency)、総購買回数 (Frequency)、総購買金額 (Monetary value) の 3 指

<sup>1</sup> 早稲田大学大学院創造理工学研究科経営システム工学専攻  
Department of Industrial and Management Systems Engineering, Graduate School of Creative Science and Engineering Waseda University, Shinjuku, Tokyo 169-8555, Japan

<sup>2</sup> 早稲田大学理工学術院  
Faculty of Science and Engineering Waseda University, Shinjuku, Tokyo 169-8555, Japan

a) khasumoto@fuji.waseda.jp

b) moto-aries@ruri.waseda.jp

c) masagoto@waseda.jp

標が広く活用されている [4], [5], [10], [15]. RFM は各指標の頭文字をとっており, Recency は直近の購買日で, 観測時点までの日数で数値化され, 一般的に買い物の頻度が少ないほど値が大きくなる. Frequency は購入の頻度で, 一般的に初回購入から観測時点までの購入回数合計である. Monetary value は購買金額で初回購入から観測時点までの購入金額合計である. 購入単価など, 平均値を用いる場合もある. これらの指標は購買履歴から容易に得られるだけでなく, 顧客行動を端的に示す指標としてアカデミックリサーチだけでなく, ビジネスの現場でも頻繁に用いられている [5].

### 2.1.2 CLV 予測モデル

CLV の研究においては, 購入数をポワソン分布で, 継続購入期間を指数分布でモデル化し, 将来の購買を予測する Pareto/NBD モデル [3] が広く知られており, それらを拡張する試み [4], [5] も多数行われている. これらのモデルは顧客の直近の購買日や購買回数である RFM 指標をベースにしており, それらは比較的容易に過去の購買履歴から取得が可能である. また RFM 指標をランダムフォレストなどの回帰モデルに説明変数として取り込み, 予測を行うこともできる. ランダムフォレスト [16] は決定木を複数用いて, 回帰・分類を行うアンサンブル手法で, 高速で精度高い手法として CLV 予測 [14] を含め, 幅広い問題に適用されている. 目的は異なるが本研究のように NMF で得られた潜在パターンをランダムフォレストで利用するという研究 [8] も報告されている. ランダムフォレストは重回帰分析と異なり, 多重共線性によりパラメータ推定が不安定になる問題が発生しないため, 相関がある特徴も同時に扱うことができる. 一方で RFM 指標のみを説明変数とした場合, パラメータ数が十分とはいえない. またプラットフォームビジネスでは一般的に店舗や商品カテゴリごとに異なる手数料率が設定されており, 同じ購入数や金額であってもプラットフォームビジネス側の収益は異なる. そのため, 店舗をまたぐ多様な顧客の購買行動や背後にあるプラットフォーム事業者と導入店舗の手数料関係が CLV に影響を与えており, それらを考慮した予測モデルが必要となる.

### 2.1.3 非負値行列因子分解

非負値行列因子分解 (NMF) は, 非負の行列データを低ランクの非負行列の積に因子分解することで, それらのデータの潜在的なパターンを抽出する手法である. 比較的単純なアルゴリズムでデータの特徴を抽出する手法として, 画像解析 [6] や文書分類 [17] など様々な問題に適用されている. また NMF のような行列分解を購買履歴に適用し, 顧客の購買パターンを抽出する研究 [18], [19] も行われており, その有効性が示されている. いま, 顧客  $I$  と購入先店舗  $J$  で構成される非負行列  $\mathbf{V} = [v_{ij}] \in \mathbb{R}_+^{I \times J}$  があり, その要素は購入数であるとする. 非負行列  $\mathbf{V}$

が顧客行列 ( $\mathbf{W} = [w_{ik}] \in \mathbb{R}_+^{I \times K}$ ) と, 購入先店舗行列 ( $\mathbf{H} = [h_{kj}] \in \mathbb{R}_+^{K \times J}$ ) の 2 つの行列に特徴数  $K$  で分解されるとすると行列は式 (2) で近似される.

$$\mathbf{V} \cong \mathbf{W}\mathbf{H} \quad (2)$$

このとき,  $\mathbf{V}$  と  $\mathbf{W}\mathbf{H}$  の差異が最小となるように式 (3) の  $F$  を目的関数として最小化する.

$$F = \|\mathbf{V} - \mathbf{W}\mathbf{H}\|^2 \quad (3)$$

NMF の距離関数にはユークリッド距離や一般化 KL ダイバージェンス, 板倉斎藤距離などが知られている [20] が, ここではユークリッド距離を用いる. また更新アルゴリズムも乗法更新ルール [7] や Projected gradient methods [21] など様々なものが提案されているが, 本研究では様々な NMF の研究で使用されており, かつ安定した収束が期待できる乗法更新ルールに基づいたアルゴリズムを用いる. 乗法更新ルールは比較的シンプルなロジックで拡張やアルゴリズムの改良が容易なことも特徴である. 式 (4), (5) はユークリッド距離での乗法更新ルールによる更新式である.

$$w_{ik} \leftarrow w_{ik} \frac{\sum_{j=1}^J v_{ij} h_{kj}}{\sum_{j=1}^J \hat{v}_{ij} h_{kj}} \quad (4)$$

$$h_{kj} \leftarrow h_{kj} \frac{\sum_{i=1}^I v_{ij} w_{ik}}{\sum_{i=1}^I \hat{v}_{ij} w_{ik}} \quad (5)$$

$\mathbf{W}$  と  $\mathbf{H}$  をランダムな非負値で初期化し, 各行列の要素を式 (4) と式 (5) を交互に更新していくことで, 局所最適解に収束することが知られている. ここで  $v_{ij}$  は入力行列  $\mathbf{V}$  の要素,  $\hat{v}_{ij}$  は  $w_{ik}$  と  $h_{kj}$  によって更新された近似値である. このとき,  $\mathbf{H}$  は  $K$  個の特徴的な購買パターンにまとめられる. 同時に  $\mathbf{W}$  は顧客ごとの特徴  $k$  への所属割合へと分解される.  $\mathbf{W}$  は顧客ごとの潜在的な購買パターンと解釈することができ, CLV 予測では回帰モデルの説明変数に取り込んで使用することができる.

## 2.2 対象事例

本研究ではプラットフォームビジネスを対象とし, プラットフォームビジネスを運営する事業者 A から提供された 2 年分の顧客データ, その購買履歴データおよび CLV を分析する. この事業者のサービスは様々な店舗で導入されており, 消費者は事前にプラットフォーム事業者でサービスの会員登録をしたうえで, サービス導入企業でショッピングを行う. 消費者が当該サービスを利用して導入店舗から商品を購入すると, 事業者には導入店舗から一定の手数料が支払われる. 消費者が継続的にこのサービスを利用して商品を購入すれば, 手数料収入も増えるため, 収益は基本的に総購買金額に比例する. 一方で手数料率は店舗や利用サービスごとに異なるため, 同じ購買金額であってもどの店舗で購入されたか, どのサービスを利用して購入し

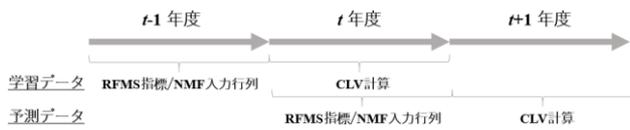


図 1 学習データ・予測データの生成

Fig. 1 Generation of training data and prediction data.

たかによって事業者の収益は異なる。また購入頻度や購入単価は消費者だけでなく、店舗や商品の特性にも依存する。事業の性質上、商品レベルの情報はない一方で複数の店舗にまたがる取引情報は存在する。そのため、顧客の嗜好に基づく購買先の多様性や時系列での購買のトレンドが将来の購買行動に影響し、CLV の増減をもたらすと考えられる。

### 2.3 RFMS 指標を用いた対象データの事前分析

本節では、目的である CLV 予測モデルを提案するに先立ち、対象とする事例のデータについて、基礎的な事例分析を行った結果について述べる。まず、本研究では回帰モデルの説明変数として RFM 指標を用いる。加えて、事前分析によりプラットフォームビジネスにおいて複数店舗での購入が、CLV に影響を与えることが判明したため、総購買店舗数 (Stores) を追加指標として導入し、RFMS 指標として利用することとした。また総購買金額 (Monetary value) は総購買回数 (Frequency) と強い相関があることから、本研究では平均購買金額を購買金額に関する指標として使用した。RFMS 指標は CLV 研究において広く利用されており、CLV に大きな影響を与える因子と考えられる。

予測の対象となるデータはある年 ( $t$  年度) のある月に初めて利用を開始した顧客データ約 8,400 人分であり、この顧客の翌年度の CLV を予測する。また学習データとして、その前年 ( $t-1$  年度) の顧客データ約 7,500 人分がある。それぞれ利用開始から 2 年分の購買履歴データがあり、図 1 に示したように 1 年目のデータから RFMS 指標と NMF の入力行列の取得、2 年目のデータから CLV の算出ができる。

図 2 は  $t$  年度の顧客の RFMS 指標と CLV で作成した相関行列をヒートマップでプロットしたものである。CLV はどの指標ともある程度相関があり、RFMS 指標が CLV に影響を与える因子であることが分かる。また Recency と Frequency にも相関がみられ、購買頻度や購買サイクルが直近の購買に関連していることを示している。さらに Frequency は Stores と相関がみられるが、これは購入数が増えれば、自然に購入先の数も増えていくことを示している。なお Recency は最終購買日から CLV の算出開始日までの日数のため、他の指標とは負の相関になる。

### 2.4 課題と本研究への展開

本研究では非負値行列因子分解 (NMF: Non-negative

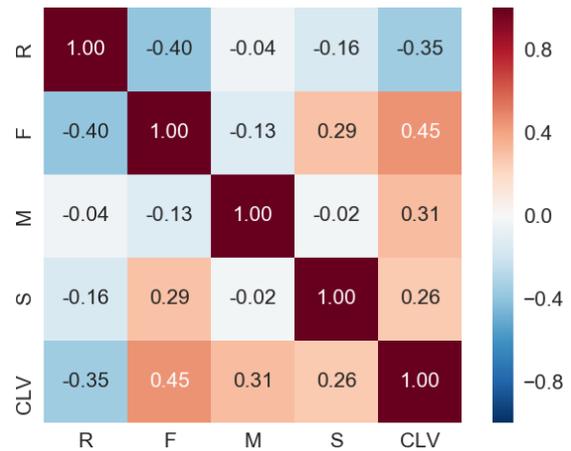


図 2 RFMS 指標と CLV の相関行列 ( $t$  年度)

Fig. 2 Correlation matrix of RFMS and CLV (Year  $t$ ).

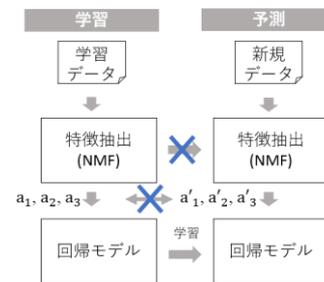


図 3 NMF と回帰モデルの組合せの問題

Fig. 3 A problem of combining NMF and a regression model.

Matrix Factorization) を用いて、顧客の潜在的な購買パターンを抽出し、それらを RFMS 指標と組み合わせ、ランダムフォレスト回帰による CLV 予測に適用する。NMF などの行列分解は消費者の購買パターン抽出においてもその有用性が示されており [18], [19], 対象事例のような店舗をまたぐ顧客の購買行動や時系列の購入傾向を抽出するのに有効であると考えられる。しかし従来の NMF で抽出した特徴を回帰モデルに組み合わせて、予測問題に適用した場合、新規の入力データに対する正しい予測ができない。それは NMF が分析対象データごとにそれぞれ異なる特徴空間を構成してしまうためである。たとえば図 3 のように学習用のデータと予測用の新規データがあり、それぞれを NMF で分析して特徴を抽出し、それを回帰モデルの説明変数として使用する 2 段階のモデルを考える。仮に学習データから  $a_1, a_2, a_3$  の 3 つ特徴が抽出されたとし、それを回帰モデルに学習させる。次に新規のデータを使って予測を行いたい、このとき、NMF で抽出される特徴  $a'_1, a'_2, a'_3$  はあくまで新規データの行列分解に基づくものであり、学習データから抽出された特徴とは連動していない。そのため、この結果を学習データで訓練した回帰モデルの入力に用いて予測を行うことはできない。学習データと新規データをあわせて特徴抽出し、学習データのみを使って、回帰モデルを学習することもできるが、データが

追加されるごとにすべてのデータで特徴抽出を行い、回帰モデルを再学習させる必要があり、購買情報のように逐次的にデータが追加される環境においては計算量の面で現実的でない。そこで本研究では学習データに対して NMF で特徴抽出した際に得られた部分行列を活用し、新規データを同じ特徴空間にマッピングする方法を提案し、この問題を解決する。

### 3. 提案モデル

#### 3.1 概要

本研究では購買パターンを考慮した顧客生涯価値の回帰予測モデルを提案する。顧客の購買パターンを回帰予測モデルで利用するのは以下 2 つの理由がある。1 つ目は予測精度の向上に役立つことである。プラットフォームビジネスでは多くの店舗が参加しており、それらを利用する顧客の購買行動は多様である。顧客の利用先店舗や時系列の購買傾向は長期的な顧客の購買傾向や事業者への収益に影響があり、CLV にも強く影響していると考えられる。そのため購買パターンを説明変数として使用することで予測精度の向上が期待できる。2 つ目は購買パターンが購買履歴から取得できるということである。CLV の予測に役立つ変数として RFMS 指標のような購買情報以外にウェブページへの訪問数や滞在時間なども考えられるが、そのようなデータは必ずしも取得できるとは限らない。一方で購買履歴は売上や在庫管理の目的で多くの企業が保有しており、容易に取得できると考えられる。

購買パターンの抽出は NMF で行い、回帰予測の説明変数として利用する。このとき、先に説明した回帰モデルとの組合せの問題が生じるが、学習データに対する NMF で得られた部分行列を活用し、この問題を解決する。このモデルでは前年度に利用を開始した顧客の実績をもとに翌年度に利用を開始した顧客の CLV を予測する。CLV は便宜上利用開始 1 年経過後、その先 1 年間で得られる収益とする。図 1 において、現在が  $t$  年度の終わりであるとする、 $t-1$  年度に利用を開始した既存顧客（以下、 $t-1$  年度の顧客）のデータを用いて、 $t$  年度に利用を開始した新規顧客（以下、 $t$  年度の顧客）によって  $t+1$  年度にもたらされる収益、つまり  $t$  年度の顧客の CLV を予測することが最終的な目的になる。

#### 3.2 定式化

提案モデルは主に学習フェーズと予測フェーズの 2 つのステップに分かれる。学習フェーズでは、まず  $t-1$  年度の既存顧客の購買履歴から NMF の入力行列を作成する。入力行列は行に顧客、列に購入店舗や購入時期などの購買情報を持つ。次にその入力行列を使って、NMF で顧客ごとの潜在的な購買パターンを抽出する。NMF の行列分解より、顧客ごとにクラスへの所属割合を示す行列が生成さ

れるが、これを購買パターンごとの特徴量ととらえることができる。具体的にはクラス数  $K$  で NMF を実行した際に得られる出力行列  $\mathbf{W}$  を  $K$  次元の購買パターンの特徴ベクトルとして利用する。さらに RFMS 指標と NMF の出力行列を説明変数、 $t-1$  年度の顧客の CLV を学習データとしてランダムフォレスト回帰による予測モデルの学習をする。

予測フェーズでは同様に  $t$  年度の顧客の購買履歴から NMF の入力行列を作成する。ここで、作成した入力行列を使って従来手法の NMF を実行すると、新たなデータで新たなクラスに分解されてしまい、 $t-1$  年度の顧客の学習結果が反映されない。 $t$  年度の顧客と  $t-1$  年度の顧客をまとめてから NMF の入力行列を作成し、クラスタリングする方法もあるが、購買情報が逐次的に追加されることを考えると、モデル全体を毎回更新することは現実的でない。NMF を逐次的に適用する手法としては Online NMF [22] があるが、新たなデータを追加する際に元のモデルも一緒に更新してしまい、今回の目的には適さない。そこで  $t-1$  年度の顧客の NMF 出力結果を  $t$  年度の顧客の NMF 実行時に利用することで  $t-1$  年度の顧客の NMF のクラスとの整合性を保ちながら、予測を行う手法を提案する。具体的には  $t-1$  年度の顧客の NMF 出力結果で分解された行列  $\mathbf{H}_{t-1}$  (クラス  $[K] \times$  購買情報  $[J])$  を固定したまま、 $t$  年度の顧客の NMF で使用し、行列  $\mathbf{W}_t$  (顧客  $[I] \times$  クラス  $[K])$  を求める。図 4 に提案モデルのイメージ図を示す。

以下で提案モデルを定式化する。学習フェーズでの NMF は、 $t-1$  年度の顧客データから作成される顧客と購買情報の行列  $\mathbf{V}_{t-1} = [v_{ij}^{t-1}] \in \mathbb{R}_+^{I \times J}$  を顧客とクラスの行列  $\mathbf{W}_{t-1} = [w_{ik}^{t-1}] \in \mathbb{R}_+^{I \times K}$  とクラスと購買情報の行列  $\mathbf{H}_{t-1} = [h_{kj}^{t-1}] \in \mathbb{R}_+^{K \times J}$  の 2 つの行列に分解する。このとき、NMF は式 (6) で表される。

$$\mathbf{V}_{t-1} \cong \mathbf{W}_{t-1} \mathbf{H}_{t-1} \quad (6)$$

次に予測フェーズでは  $t$  年度の行列  $\mathbf{V}_t$  を追加することを考えるが、 $\mathbf{H}_{t-1}$  のもとに  $t-1$  年度の既存顧客データと  $t$  年度の新規顧客データが分解されると考える。このとき、 $\mathbf{W}_{t-1}$  と  $\mathbf{H}_{t-1}$  は固定とし、式 (7) のモデルを考える。

$$\begin{pmatrix} \mathbf{V}_{t-1} \\ \mathbf{V}_t \end{pmatrix} \cong \begin{pmatrix} \mathbf{W}_{t-1} \\ \mathbf{W}_t \end{pmatrix} \mathbf{H}_{t-1} \quad (7)$$

予測フェーズでは  $\mathbf{V}_t$  と  $\mathbf{H}_{t-1}$  が与えられているため、最小二乗法により  $\mathbf{W}_t$  を求めることもできるが、この方法では非負性が保証できない。そこで本研究では、非負性を保証するため、式 (8) を乗法更新ルールを使って解くことで解を求める。この問題では  $\mathbf{H}_{t-1}$  は固定されているため、 $\mathbf{W}_t$  をランダムな非負値で初期化したのち、式 (9) と式 (10) を交互に更新し、 $\mathbf{W}_t$  を求める。

$$F = \|\mathbf{V}_t - \mathbf{W}_t \mathbf{H}_{t-1}\|^2 \quad (8)$$

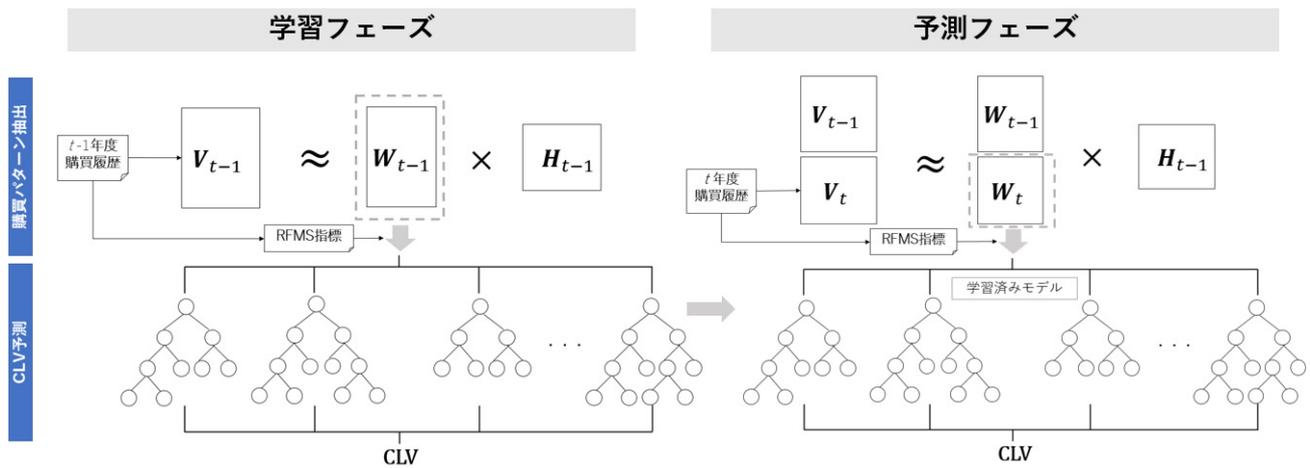


図 4 提案モデルのイメージ図

Fig. 4 The Image diagram of the proposed model.

$$w_{ik}^t \leftarrow w_{ik}^t \frac{\sum_{j=1}^J v_{ij}^t h_{kj}^{t-1}}{\sum_{j=1}^J \hat{v}_{ij}^t h_{kj}^{t-1}} \quad (9)$$

$$\hat{v}_{ij}^t \leftarrow \sum_{k=1}^K w_{ik}^t h_{kj}^{t-1} \quad (10)$$

上記より  $t-1$  年度の顧客データからは  $W_{t-1}$  が、 $t$  年度の顧客データからは  $W_t$  が得られる。学習フェーズでは  $t-1$  年度の顧客の RFMS 指標と  $W_{t-1}$  をランダムフォレスト回帰の説明変数として使用し、予測フェーズでは  $t$  年度の顧客の RFMS 指標と  $W_t$  を説明変数とする。

### 3.3 提案モデルによる CLV 予測アルゴリズム

以下に提案モデル全体のアルゴリズムを示す。

#### 【CLV 予測アルゴリズム】

[学習フェーズ]

1.  $t-1$  年度の顧客データの NMF (学習データ)

**Step 1)**  $w_{ik}^{t-1}$  と  $h_{kj}^{t-1}$  をランダムな非負値で初期化する。

**Step 2)** 式 (4) より、 $w_{ik}^{t-1}$  を求める。

**Step 3)**  $w_{ik}^{t-1}$  を式 (5) に代入し、 $h_{kj}^{t-1}$  を計算する。

**Step 4)**  $F$  が収束するまで、Step 2), 3) を繰り返す。

2. CLV 予測のランダムフォレスト回帰モデルの学習

**Step 5)**  $t-1$  年度の顧客の RFMS 指標と  $w_{ik}^{t-1}$  を説明変数、翌年の CLV を目的変数とする学習データを作成する。

**Step 6)** Step 5) で作成した学習データを用いて、ランダムフォレスト回帰のモデルを学習する。

[予測フェーズ]

3.  $t$  年度の顧客データの NMF (予測データ)

**Step 7)**  $w_{ik}^t$  をランダムな非負値で初期化する。

**Step 8)** 式 (9) より、 $w_{ik}^t$  を求める。

**Step 9)**  $w_{ik}^t$  を式 (10) に代入し、 $\hat{v}_{ij}^t$  を計算する。

**Step 10)**  $F$  が収束するまで、Step 8), 9) を繰り返す。

### 4. CLV 予測

**Step 11)**  $t$  年度の RFMS 指標と  $w_{ik}^t$  から予測データを作成する。

**Step 12)** Step 6) で学習したモデルに予測データを入力し、ランダムフォレスト回帰による CLV 予測を行う。

## 4. 実データによる分析と顧客生涯価値予測

提案モデルによる NMF での購買パターンの抽出、ランダムフォレスト回帰による顧客生涯価値予測の予測精度について、プラットフォーム事業者 A から提供された実データを用いて検証した。

### 4.1 実験データ

実験データは  $t-1$  年度における新規顧客データ約 7,500 人分の初回購入後 1 年間の購買情報、翌 1 年間の購買からの CLV、同様に  $t$  年度の新規顧客データ約 8,400 人分と購買情報、CLV である。 $t-1$  年度のデータは学習データとして使用し、 $t$  年度のデータは予測データとして、CLV の予測精度評価に使用する。提供された購買情報は、購入先と時系列の購入傾向を金額と数量で 1 つの行列にまとめ、NMF でパターン抽出する。ただし購入先は店舗数が膨大であり、行列が非常にスパースになっているため、そのまま統計モデルに取り込むと推定精度が大幅に劣化してしまう。そこで店舗が取り扱っている商品 (ファッション、食品など) によって店舗カテゴリを構成し、顧客がどの店舗カテゴリの店舗から購入しているのかを行列表現して NMF で行列分解する。このとき購入先を店舗カテゴリに集約することで情報量が失われるが、顧客の購買行動における購入先の多様性は、顧客が嗜好する商品の種類、つまり店舗カテゴリで大きく決まるため、本実験のように購買パターンの抽出に着目した場合、情報量の損失の影響は少ないと考えられる。また、店舗カテゴリを用いた場合でも、各顧客

顧客ID	購入数								購入金額							
	購入店舗カテゴリ				購入月				購入店舗カテゴリ				購入月			
	ソフトウェア	自動車用品	...	ファッション	1か月目	2か月目	...	12か月目	ソフトウェア	自動車用品	...	ファッション	1か月目	2か月目	...	12か月目
A001	1				1			1	5000				300			600
A002					1								7800			
A003	1			2	1				7000			7500	580			
A004		1			2	1				30000			7000	500		
A005					1								2900			
A006					4								6000			
A007				4	2							24000	20000			
A008	3			3	3				9900			38000	15000			

図 5 入力行列の一例

Fig. 5 A sample of input matrix.

表 1  $t$  年度の顧客の CLV 予測結果

Table 1 The result of predicting CLV for Year  $t$  customers.

説明変数	学習データ( $t-1$ 年度)		予測データ( $t$ 年度)	
	平均二乗誤差(MSE)	決定係数(R2)	平均二乗誤差(MSE)	決定係数(R2)
RFMS指標(4)	1.2816	0.6765	2.5563	0.5001
RFMS指標(4)+NMF入力行列(118)	1.1852	0.6961	2.5580	0.4918
RFMS指標(4)+NMF出力結果(3)	1.0979	0.7228	2.4997	0.5110
RFMS指標(4)+NMF出力結果(4)	<b>1.0717</b>	<b>0.7294</b>	2.4974	0.5114
RFMS指標(4)+NMF出力結果(5)	1.0854	0.7259	<b>2.4925</b>	<b>0.5123</b>
RFMS指標(4)+NMF出力結果(6)	1.0786	0.7275	2.5321	0.5044
RFMS指標(4)+NMF出力結果(7)	1.0970	0.7229	2.5681	0.4973

の購買履歴は一部の店舗カテゴリに偏っており、依然としてスパースなデータである。図 5 に入力行列の例を示す。

購買情報は顧客ごとに時系列情報を月単位、購入先を購入店舗カテゴリで集計し、入力行列を作成した。購入月は初回購入月を 1 か月目とし、12 か月目までの 12 列である。また購入先の店舗カテゴリ数は 47 である。上記のデータに対して、購入回数、金額を要素とする行列を作成し、学習データの NMF の入力行列は  $7,508 \times 118$ 。同じく予測データの入力行列は  $8,419 \times 118$  となる。また RFMS 指標はそれぞれの年度で集計し使用した。

## 4.2 実験結果

### 4.2.1 CLV 予測結果

提案手法を用いて CLV 予測を行った結果を表 1 に示す。クラス数  $K$  は 1~30 の範囲で事前分析を行い、NMF 実行時の残差と実行結果の解釈容易性から  $K$  の範囲を 3~7 に絞らなうえで実験を行った。評価値は平均二乗誤差 (MSE) と決定係数 (R2) を使用し、比較手法として RFMS 指標のみを使用した場合、RFMS 指標と NMF の入力行列 (購入月行列 12 列、店舗行列 47 列が購入数と金額で 118 変数) を使用した場合の実験も行った。まず RFMS 指標のみを使用した場合と比較することで、NMF によって購買パターンを抽出し、CLV 予測に活用することが予測精度の向上につながることを示す。しかし購買パターンの抽出に使用する NMF の入力データには、時系列データなど RFMS 指標にはない情報が含まれている。そこで入力情報

表 2 クラス数  $K = 5$  での特徴重要度

Table 2 Feature importance at  $K = 5$ .

#	特徴	重要度
1	M	0.2811
2	F	0.2217
3	R	0.2123
4	S	0.0923
5	$k=3$	0.0520
6	$k=5$	0.0379
7	$k=1$	0.0366
8	$k=2$	0.0349
9	$k=4$	0.0312

表 3 クラス別の CLV と出現数 ( $t-1$  年度の顧客)

Table 3 CLV and number of occurrences of classes (Year  $t-1$  customers).

Class( $k$ )	1	2	3	4	5
出現数	1238	3344	1245	963	718
出現数(%)	16.5%	44.5%	16.6%	12.8%	9.6%
平均 CLV	0.73	0.94	1.32	0.79	1.47

が同じ場合でもより高い精度での予測が可能であることを示すために RFMS 指標と NMF の入力行列を使用した場合での比較も行う。なお表 1 中の () 内の数字は説明変数の数を表す。学習データでは MSE, R2 とともに提案手法でクラス数  $K = 4$  のときに最も良い結果を示した。一方で、予測データでは提案手法で  $K = 5$  のときに MSE, R2 が最

表 4 上位 5 購入店舗カテゴリ出現順位  
Table 4 Top 5 Categories of purchased stores.

クラス	出現数	出現順位 (購入数)					出現順位 (購入金額)				
		1	2	3	4	5	1	2	3	4	5
$k=5$	718	サービス2	非耐久消費財1	メディア1	ホビー1	非耐久消費財2	サービス2	メディア1	非耐久消費財1	トラベル1	サービス1
$k=3$	1245	その他1	耐久消費財1	ホビー2	耐久消費財2	ファッション1	その他1	サービス3	デジタルメディア1	非耐久消費財3	耐久消費財1
$k=2$	3344	サービス3	サービス4	耐久消費財2	デジタルメディア1	メディア1	サービス3	耐久消費財2	サービス1	ファッション1	サービス4
$k=4$	963	デジタルメディア2	サービス1	その他1	デジタルメディア1	メディア1	デジタルメディア2	デジタルメディア1	メディア1	耐久消費財1	ホビー3
$k=1$	1238	デジタルメディア3	サービス1	デジタルメディア1	デジタルメディア2	サービス4	デジタルメディア3	デジタルメディア1	サービス1	ホビー4	サービス4

も高く、予測精度が良かった。予測データにおいて最適なクラス数が増えているのは、当該企業が成長企業であり、前年の学習データよりも翌年の予測データの方が顧客ごとの購入が増え、購買が多様になっていることが考えられる。ただし新しい購買パターンが現れたわけではなく、潜在的に存在していた購買パターンがより強く観測されるようになったと考えられる。どちらの場合も  $K = 7$  の場合を除いて、提案手法が比較手法よりも高い精度を示した。

クラス数  $K = 5$  で予測を行ったときの特徴重要度を表 2 に示す。RFMS 指標とともに CLV の平均値が高い NMF 第 3 特徴量 ( $k = 3$ ) や NMF 第 5 特徴量 ( $k = 5$ ) が上位になっていることが分かった。

4.2.2 NMF による購買パターン抽出 (学習フェーズ)

最終的な CLV の予測精度が最も高くなった  $K = 5$  のときのクラス別の CLV と出現数を表 3 に示す。ただし NMF の結果は顧客ごとの特徴  $k$  への所属割合を返すため、出現数は 1 に基準化した所属割合の合計、平均 CLV は顧客ごとの CLV を所属割合に応じて案分し、算出した。式 (11), (12) に計算式を示す。ここで  $CLV_i$  は顧客  $i$  ごとの CLV を表す。

$$\text{出現数} = \sum_{i=1}^I \frac{w_{ik}^{t-1}}{\sum_{k=1}^K w_{ik}^{t-1}} \quad (11)$$

$$\text{平均 CLV} = \frac{\sum_{i=1}^I \left( CLV_i \times \frac{w_{ik}^{t-1}}{\sum_{k=1}^K w_{ik}^{t-1}} \right)}{\sum_{i=1}^I \sum_{k=1}^K \frac{w_{ik}^{t-1}}{w_{ik}^{t-1}}} \quad (12)$$

さらに購入店舗カテゴリの出現順位 (上位 5 カテゴリ) を表 4 に、購入月の分解結果を時系列でグラフ化したものを図 6 に示した。

購入月の傾向を見ると購入数、購入金額ともに利用初期に購入が多く後半にかけて減っていく場合 ( $k = 1, 2, 4$ ) は CLV が低く、逆に購入数や金額が減少せず、継続的な購入がみられる場合 ( $k = 3, 5$ ) は CLV が高くなっていることが分かる。また購入金額が 12 カ月目において高い場合は CLV も高い傾向にあり、Recency がその後の CLV にポジティブな影響を与えていることを示している。CLV 上位クラスの利用先傾向としては非耐久消費財やメディア商材など購入サイクルが比較的短い商材を扱う店舗カテゴリや、耐久消費財やトラベル、ホビーなど比較的単価が高い店舗カテゴリが多く観測された。一方で CLV 下位クラスの傾向としてデジタルメディアなど単価の低いクラスが出現しており、購入頻度や平均購入金額による CLV への影

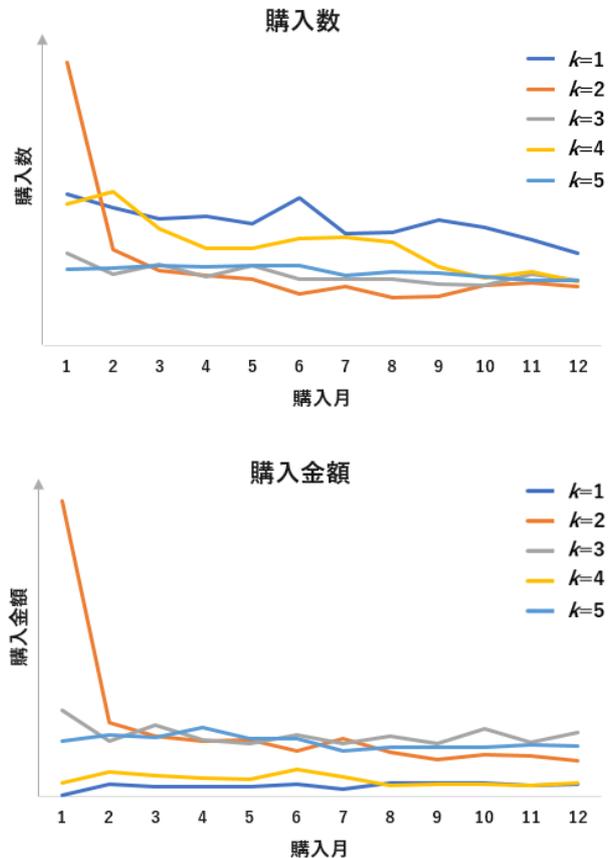


図 6 購入月の分解結果  
Fig. 6 Result of decomposition for purchase month.

響が観測された。購入数と購入金額でカテゴリの出現順位に大きな差異はみられなかったが、サービス 3 やホビー 2 のように同じクラスで出現しながら、どちらか一方のみで観測されるケースもあった。これらは単価による影響と考えられる。

4.2.3 NMF による購買パターン抽出 (予測フェーズ)

$K = 5$  で NMF を実行した際のクラス別の CLV と出現数を表 5 に示す。表 3 の  $t - 1$  年度の結果と同様に  $k = 3, 5$  のクラスで CLV が高く、 $k = 1, 4$  が低くなっている。出現数のクラスごとの割合も  $k = 5$  が最も少なく、 $k = 2$  が最も多い点も同じであった。 $t - 1$  年度の顧客データから得られた購買パターンと CLV の値が、 $t$  年度の顧客データを使った NMF でも同じクラスで引き継がれていることが分かる。 $t$  年度の CLV は  $t - 1$  年度と比べて全体的に高くなっているが、これは当該企業自体が成長しており、顧客自身の購入数・購入金額が増えているためである。一方で

表 5 クラス別の CLV と出現数 ( $t$  年度の顧客)

Table 5 CLV and number of occurrences of classes (Year  $t$  customers).

Class( $k$ )	1	2	3	4	5
出現数	1199	3468	1752	1385	615
出現数(%)	14.2%	41.2%	20.8%	16.4%	7.3%
平均 CLV	0.90	0.98	1.58	0.79	1.54

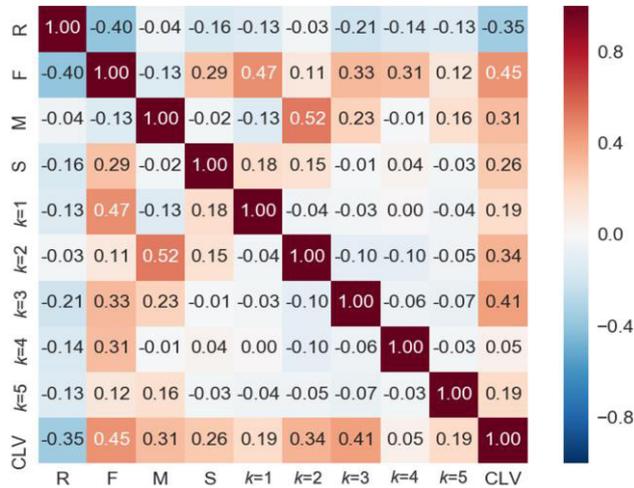


図 7 RFMS 指標,  $W_t$  計算結果, CLV の相関行列 ( $t$  年度)

Fig. 7 Correlation matrix of RFMS index,  $W_t$ , CLV (Year  $t$ ).

クラス間の傾向は引き継がれており、年度が変わっても購買パターンは大きく変化していないことが推測される。

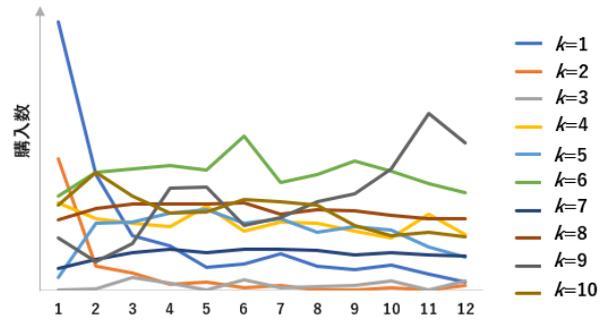
$W_t$  の計算結果と RFMS 指標, CLV との相関行列をプロットしたものを図 7 に示す。出力クラスと CLV との関係を見ると,  $k=2, k=3$  で最も強い相関がみられた。また同時に  $k=2$  は平均購買金額 (Monetary value) と相関がみられた。

### 5. 考察

最終的な予測結果は、クラス数  $K=5$  のときの提案手法が最も良い精度を示した。すなわち, NMF により潜在的な購買パターンを抽出し, CLV 予測の説明変数として追加することで, 単純に RFMS 指標を使うよりも予測精度が高まったといえる。一方で  $K=7$  の場合には RFMS 指標のみを使用した場合の精度を下回っており, クラス数の選択が予測精度に大きく影響することを示している。 $t-1$  年度の既存顧客と  $t$  年度の新規顧客の NMF の結果を比較すると, 出現割合や CLV の平均値で同じような傾向がみられ, 学習データでのクラスタリングの結果が予測データの NMF でも適切に反映されていることが分かる。

NMF による購買パターン抽出ではクラス別の平均 CLV から時系列の購入傾向や購入先により, CLV に差異ができることが確認できた。一方で時系列の購入傾向は図 6 に示したように月ごとの変動が少なく, 各クラスとも比較的安定していた。これには 2 つの要因が考えられる。1 つ目

購入数( $K=10$ )



購入金額( $K=10$ )

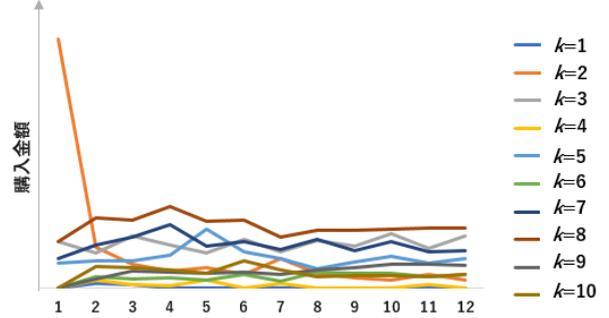


図 8 購入月の分解結果 ( $K=10$ )

Fig. 8 Result of decomposition for purchase month ( $K=10$ ).

は単純に予測精度を優先しクラス数を抑えたことで抽出されるパターンが限定されたことである。たとえばクラス数  $K=10$  で購入月の時系列分析を行うと図 8 のように, 月ごとの違いや購買パターンの多様性がより多くなることが確認できる。2 つ目は NMF の入力行列に起因するものである。今回の実験では時系列の購入月データ (12 次元) と購入店舗カテゴリ (47 次元) のデータがあり, それらをまとめて NMF でクラスタリングしている。購入月データは次元数が相対的に少ないため, 結果的に時系列データよりも次元数の多い購入店舗カテゴリのデータがクラスタリングで優先され, 時系列データの多様性が失われたと考えられる。一方で細かく購買パターンを分けることが必ずしもより精度の高い予測につながるわけではない。今回のケースでは購入先店舗カテゴリに強く影響を受ける平均購買金額 (Monetary value) は特徴重要度で 1 位となっており, 購入先の潜在的な購買パターンを適切に抽出することで, 予測精度の改善につながったと考えられる。相関行列のヒートマップではいくつかのクラスと CLV に一定の相関がみられた。これは抽出された購買パターンの CLV への影響を示している。一方で特徴重要度では RFMS 指標がすべて上位になっており, 本研究で抽出した購買パターンではそれを上回ることはできなかった。潜在的な顧客の購買パターンが顧客の購買行動に影響を与えていることは分かったが, 購買頻度や金額のような直接的な指標である RFMS 指標が CLV により大きな影響を与えていると考えられる。NMF の特徴量の中では, 第 3 特徴量 ( $k=3$ ) と第 5 特徴

量 ( $k=5$ ) が上位であった。これらの特徴量は  $t-1$  年度も  $t$  年度も、出現率 (%) は合計で 30% に満たないが、平均 CLV は他の NMF 特徴量比べて突出して高い。すなわち、これらの特徴量は情報量が大きく、これらで高い値が観測されれば CLV も大きくなる傾向があるため、結果的に特徴重要度が高くなったと考えられる。図 7 の相関行列によると NMF 第 3 特徴量 ( $k=3$ ) は、CLV との相関も強く、それを裏付けている。一方で NMF 第 5 特徴量 ( $k=5$ ) は CLV との相関は高くはなかったが、これは NMF で抽出された特徴と CLV の関係が非線形、かつパラメータ間の相互作用も含む複雑な関係にあるためと推測される。

本実験では、購買データの背後にある顧客の購買パターンに着目し、NMF による抽出を行ったが、異なるデータセットに提案手法を適用した場合でも、時系列データの予測精度の向上が期待できると考えられる。本研究で対象としている問題は、様々な店舗カテゴリで購入を行う新規顧客の購買履歴データから翌年の CLV を予測するという問題設定であり、1 人 1 人の顧客の購買履歴データがスパースであることを想定している。クレジットカードやポイントカードの利用履歴データなどにおいても、利用される店舗数や店舗カテゴリが多く、1 人 1 人の顧客の購買履歴データがスパースであることは、一般的な状況として仮定しても差し支えないと考えられる。加えて、ユーザの嗜好は多様であり、行動基準や価値観の異なるユーザが混ざり合っており、全体の母集団を構成しているという状況も一般的である。NMF では入力行列にあるデータの潜在的なパターンを次元圧縮することで特徴抽出するが、これらは、スパースな履歴データの背後に存在する潜在的な特徴を抽出しており、いわばスパースなデータを凝縮した特徴量を与えている。加えて、これらの特徴量によって、各ユーザの嗜好の多様性を表現することも可能である。以上のことから、これらの潜在的な特徴が目的とする関数と関連していると仮定できる対象問題の場合、本提案手法を適用することで同様に予測精度の改善効果を期待できると考えられる。

今回は CLV の計算期間を 1 年とし、新規顧客の翌年の CLV を予測するという比較的短期間の予測モデルを扱ったため、購買履歴とその後の購買行動の関連性は比較的高かったと考えられる。これが時系列上さらに長期にわたった場合、顧客の購買行動はより多様になり、予測がさらに難しくなることが考えられる。一方で長期にわたる顧客行動であっても、当初 1 年間の購買パターンになんらかの潜在的嗜好のサインが現れると考えられる。たとえば、月額動画サービスやオンライン英会話のようなサブスクリプション型の店舗カテゴリでの購入は長期的な利用の可能性を示すサインと想定され、長期的な顧客行動の予測を容易にすると考えられる。そのため、時系列の観測期間がより長くなった場合、多様性の増加により全体的な予測精度は低下するが、市場が比較的定常であれば、提案手法による

予測精度の向上は引き続き有効であると考えられる。一方で、本研究で対象としたプラットフォーム事業のビジネスを取り巻く環境は、新技術の発明や新規事業者の参入などの様々な影響を受け、非定常に変化することが考えられる。より長期間のデータが利用可能となった場合の長期予測モデルの構築においては、ビジネス環境要因を取り込む方法や非定常な時系列モデルの構築など、慎重な検討が必要と考えられる。

## 6. まとめと今後の課題

本研究では、非負値行列因子分解を用いて、顧客の購買パターンを抽出し、顧客生涯価値の予測に適用するモデルを提案した。またプラットフォームビジネスにおける実データを用いた実験を行い、その有効性を示した。さらに RFMS 指標、NMF で抽出された購買パターン、CLV の相関行列をヒートマップで可視化し、特徴重要度とあわせて確認することで、当該事業者における顧客の重要指標や購買行動の分析を行った。RFMS 指標の CLV への影響度を確認するとともに、潜在的な購買パターンを補助的に用いることで CLV の予測精度を向上させられることが確認できた。

今回は NMF で時系列データと購入先のデータを同時にクラスタリングし、購買パターンとして抽出したため、それぞれの CLV への影響が分かりにくくなってしまったが、今後は NMF や他のクラスタリング手法を活用することで、より高い精度の購買パターン抽出と CLV 予測ができる可能性がある。機械学習がますます社会に浸透するにつれて、ビジネスの意思決定に利用する場面は今後も増えてくると考えられる。その際に重要になるのは、単純な自動化やブラックボックス化された仕組みの中で解を出すだけでなく、解釈のための可視化が重要になるであろう。単純な予測精度の向上だけでなく、ビジネスのアカウントビリティを満す機械学習モデルの検討は今後の課題である。

謝辞 本研究にあたり、多くのご助言をいただいた上智大学の山下遥先生、後藤研究室の方々、また貴重なデータを提供いただいたプラットフォーム事業者 A の皆様に深く感謝いたします。

## 参考文献

- [1] 富士通総研・早稲田大学ビジネススクール根来研究室 (編著): プラットフォームビジネス最前線, pp.2–13, 翔泳社 (2013).
- [2] Gupta, S. and Lehmann, R.D.: Managing customers as investments, Wharton School Publishing (2005). スカイライトコンサルティング (訳): 顧客投資マネジメント, 英治出版 (2005).
- [3] Schmittlein, C.D., Morrison, G.D. and Colombo, R.: Counting your customers: Who-are they and what will they do next?, *Management Science*, Vol.33, No.1, pp.1–24 (1987).

- [4] Fader, S.P., Hardie, G.S.B. and Lee, L.K.: RFM and CLV: Using iso-value curves for customer base analysis, *Journal of marketing research*, Vol.42, No.4, pp.415-430 (2005).
- [5] 阿部 誠: RFM データを用いた顧客生涯価値の算出: 既存顧客の維持介入と新規顧客の獲得, *マーケティングジャーナル*, Vol.34, No.1, pp.73-90 (2014).
- [6] Lee, D.D. and Seung, H.S.: Learning the parts of objects by non-negative matrix factorization, *Nature*, Vol.401, pp.788-791 (1999).
- [7] Lee, D.D. and Seung, H.S.: Algorithms for non-negative matrix factorization, *Proc. NIPS'01*, pp.556-562 (2001).
- [8] 津村拓海, 齊藤史哲, 石津昌平: ランダムフォレストを用いた特許に関する文書データからの技術適用領域に関する知識抽出, *日本経営工学会論文誌*, Vol.68, No.3, pp.161-170 (2017).
- [9] 柘植 覚, 獅々堀正幹, 北 研二: Non-negative Matrix Factorization を用いた情報検索, *情報処理学会研究報告*, 2000-NL-142, Vol.2001, No.20, pp.1-6 (2001).
- [10] Khajvand, M., Zolfanghar, K., Ashhoori, S., et al.: Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study, *Procedia Computer Science*, Vol.3, pp.57-63 (2011).
- [11] Berger, D.P. and Nasr, I.N.: Customer lifetime value: Marketing models and applications, *Journal of Interactive Marketing*, Vol.12, No.1, pp.17-30 (1998).
- [12] Gupta, S., Hanssens, D., Hardie, B., et al.: Modeling Customer Lifetime Value, *Journal of Service Research*, Vol.9, No.2, pp.139-155 (2006).
- [13] Chamberlain, P.B., Cardoso, A., Bryan Liu, C.H., et al.: Customer lifetime value prediction using embeddings, *Proc. ACM SIGKDD* (2017).
- [14] Vanderveld, A., Pandey, A., Han, A., et al.: An engagement-based customer lifetime value system for e-commerce, *Proc. ACM SIGKDD* (2017).
- [15] Cheng, C. and Chen, Y.: Classifying the segmentation of customer value via RFM model and RS theory, *Expert Systems with Applications*, Vol.36, No.3, pp.4176-4184 (2009).
- [16] Breiman, L.: Random Forests, *Machine Learning*, Vol.45, No.1, pp.5-32 (2001).
- [17] 澤田 宏: 非負値行列因子分解 NMF の基礎とデータ/信号解析への応用, *電子情報通信学会誌*, Vol.95, No.9, pp.829-833 (2012).
- [18] 幸島匡宏, 松林達史, 澤田 宏: 属性情報を考慮した消費者行動パターン抽出のための非負値多重行列因子分解法, *人工知能学会論文誌*, Vol.30, No.6, pp.745-754 (2015).
- [19] 松林達史, 幸島匡宏, 林 亜紀ほか: 非負値テンソル因子分解を用いた購買行動におけるブランド選択分析, *人工知能学会論文誌*, Vol.30, No.6, pp.713-720 (2015).
- [20] 亀岡弘和: 負値行列因子分解の音響信号処理への応用 (<小特集> 近年の音響信号処理における数理科学の進展), *日本音響学会誌*, Vol.68, No.11, pp.559-565 (2012).
- [21] Lin, C.-J.: Projected gradient methods for nonnegative matrix factorization, *Neural Comput.*, Vol.19, No.10, pp.2756-2779 (2007).
- [22] Cao, B., Shen, D., Sun, J.-T., et al.: Detect and Track Latent Factors with Online Nonnegative Matrix Factorization, *IJCAI*, Vol.7, pp.2689-2694 (2007).



蓮本 恭輔 (学生会員)

1998 年東北大学地球工学科卒業。2010 年コーネル大学ジョンソン経営大学院修了 (MBA)。2017 年早稲田大学大学院創造理工学研究科博士後期課程。株式会社 SAP ジャパン, アマゾンジャパン株式会社を経て, 現在, PayPal Pte. Ltd. にてマーケティング業務に従事。機械学習のマーケティングへの応用に興味を持つ。



雲居 玄道 (学生会員)

2008 年早稲田大学理工経営システム工学部卒業, 2008 年早稲田大学理工学術院総合研究所嘱託研究員。2015 年浄土真宗本願寺派総合研究所研究助手。2017 年早稲田大学大学院創造理工学研究科博士後期課程, 現在に至る。情報数理応用・テキストマイニングの研究に従事。



後藤 正幸 (正会員)

1994 年武蔵工業大学大学院修士課程修了。2000 年早稲田大学大学院理工学博士課程修了。博士 (工学)。1997 年早稲田大学理工助手。2000 年東京大学大学院工学研究科助手。2002 年武蔵工業大学環境情報学部助教授。2008 年早稲田大学創造理工学部経営システム工学科准教授。2011 年同大学教授。情報数理応用とデータサイエンスの研究に従事。著書に『入門パターン認識と機械学習』, コロナ社 (2014), 『ビジネス統計~統計基礎とエクセル分析』, オデッセイコミュニケーションズ (2015) 等。IEEE, 電子情報通信学会, 人工知能学会, 日本経営工学会, 経営情報学会等会員。