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

Brand-Choice Analysis using Non-negative Tensor Factorization

松林 達史
Tatsushi Matsubayashi

日本電信電話株式会社 NTT サービスエボリューション研究所
NTT Service Evolution Laboratories
matsubayashi.tatsushi@lab.ntt.co.jp, <http://kecl.cslab.ntt.co.jp/~tatsushi/>

幸島 匡宏
Masahiro Kohjima

(同 上)
kohjima.masahiro@lab.ntt.co.jp

林 亜紀
Aki Hayashi

(同 上)
hayashi.aki@lab.ntt.co.jp

澤田 宏
Hiroshi Sawada

(同 上)
hiroshi.sawada@lab.ntt.co.jp

keywords: clustering, brand choice, non-negative tensor factorization, NTF

Summary

In marketing science field, modeling of purchase behavior and analysis of brand choice are important research tasks. This paper presents a method that enables such analysis by time-series pattern extraction based on Non-negative Tensor Factorization (NTF). The development of the scanning devices and electronic payments (e.g. online shopping, mobile-phone wallet and electronic money) has led to the accumulation of more detailed POS data including the information about purchase shop, amount of payment, time, location and so on and it brings possibilities for more deep understanding of purchasing behaviors. On the other hand, due to the increase of the number of attributes, it is still difficult to effectively and efficiently handle large feature quantities. In this paper, we consider feature quantities as high-order tensor. Then, using NTF for simultaneous decomposition of multiple attributes, we show analytic effectiveness of pattern factorization for real Beer Item/Brand purchase data. By applying NTF considering three axes: USER-ID $\times$ TIME-STAMP $\times$ ITEM-ID, we find several temporal tendencies depending on the season. In addition, by focusing on the purchase-pattern correlations between beer items and brands, we find that the tendencies of brand choice strategies appear on the graph drawing results.

1. はじめに

マーケティングサイエンスの分野では、ユーザの消費行動分析においてブランド選択分析は重要な研究課題の一つである。例えば、商品販売を行っているクライアント企業では、自社ブランドから他社ブランドへの流出防止や、自社ブランドへの取り込みなど、商品開発戦略などにおいてブランド選択分析を重要視してきた。従来技術では、ロジスティック回帰分析 [Guadagni 83] など、ブランド間の遷移を回帰的に解く手法 [Winer 86] が主流であり、各ブランドの購買時系列データに対し、購買遷移を回帰的に分析を行ってきた [hardie 93, Revelt 98, Andrews 02]。また、階層ベイズモデルなどを用いた手法 [Allenby 98, Arora 98] も利用されている。しかしながら近年では、スキャン・パネルデータのように、ユーザレベルでの詳細なログ収集が行われるようになり、実際の購買遷移をユーザ単位で分析することが可能となってきた [Erdem 96, 近藤 99, 阿部 05]。それゆえ、ユーザ単位での購買傾

向など、より詳細な分析手法が必要となってきた。

スキャン・パネルデータでは、POS データに代表される購買履歴などの構造化されたデータは、購買時刻、購買店舗、アイテム情報等を用いてデータを行列として表現することが可能である。これらの行列表現されたデータの中に潜在するパターンの抽出方法として、非負値行列因子分解 (Non-negative Matrix Factorization, NMF) [Lee 01, 澤田 12] と呼ばれる手法の有用性がこれまで示されてきている [Takeuchi 13a, 幸島 13]。NMF の適用により入力となる行列データは、低次の行列積へと因子分解され、この低次行列がそれぞれのランクに対応する行要素と列要素に特徴的なパターンが抽出される。例えば、アイテム購入の時系列で因子分解を行った場合、季節性による購買パターンや、セールや広告効果による特徴パターンの抽出が期待できる。ここで購買傾向の潜在的要因の例を図 1 に示す。

従来のような単純なクラスタリング分類 (例, K-means clustering) では、購買データベクトルの類似度を抽出し、

図1 購買傾向の潜在的な要因例

図2 クラスタリング分類によるパターン抽出例

図3 因子分解法によるパターン抽出例

類似したパターンの集合から特徴的なパターンを抽出する。しかしながら、分布の類似パターンのみを抽出した場合では、複数の要因から形成されるパターンを抽出することができず、また、ログの少ないパターンは全て同一クラスタとして分類されてしまう。例えば図2に示すように、A-Fの6個の購買パターンを分析した時に、クラスタリング手法では図のように2個に分類される。一方因子分解法では、複数パターンから潜在的な特徴パターンを抽出することが可能であり、要因を切り分けた分析が可能となる。例えば図3に示すように、A-Fの6個の購買パターンを分析した時に、3つの潜在的な特徴パターンが抽出することが可能であり、これは図1で示した3つの要因が抽出可能となることを示す。

しかしながら、図3で示した因子分解を行うには、単純な時間軸方向のみの分解では、商品の購買傾向や、ユーザの嗜好を抽出することは難しい。つまり、時系列情報の他に個々の商品や、ユーザ単位での因子分解を行い、購買傾向の抽出を行う必要がある。NMFではデータ構造を行列として扱うために(ユーザ)×(時間)や(商品)×(時間)といった、二つの要素軸での分析しか実

施することができない。そこで本研究では、NMFの高次元因子分解拡張技術である、非負値テンソル因子分解(Non-negative Tensor Factorization, NTF) [Shashua 05]を用いて購買傾向の高次パターン分析を試みる。NTFではデータ構造を3次以上のテンソルとして表現することによって(ユーザ)×(時間)×(商品)といった粒度の因子分解が可能になり、ユーザ単位の購買傾向と時系列的特徴と個々の商品購買傾向を、合わせて因子分解を行うことが可能になる。本稿ではNTFを用いて、ユーザ、時間、商品のそれぞれに対しての因子行列を抽出し、時間と商品に対する特徴抽出の分析を行う。さらに、商品のランク寄与度を用いて購買傾向類似度を抽出し、得られたグラフデータから商品のグラフ可視化を行い、商品間のブランド嗜好分析を実施した。

以下、2章では本研究で利用したスキャンパネルデータに関する事前分析を示し、3章ではNTFに関する技術説明を示す。4章ではNTFによる解析結果とその分析に関する報告を行い、5章ではさらにグラフ可視化を用いたアイテム分析に関する考察を示し、6章にてまとめを示す。

2. 分析実験データ

本研究では、株式会社インテージのスキャンパネルデータ(SCI)を用いて分析を行った(以降データと略す)。データの収集期間は2013年1月1日から12月31日までの1年間分のデータを利用。データの商品情報には「商品小分類」が付与され、アイテムのカテゴリを判別することが可能であり、本研究では「ビール」に分類されたカテゴリに絞った。ビールの購買は年間を通して頻度高く定期的に購入され、かつマーケティング分析の分野でもいくつか分析が行われている[大石 06, 土井 14]。そのため、ブランド選択問題としてビール購買に絞り分析を行った。

本データでは、全ビール購買ログ数(84,587回)のうちの約9割の77,926回が、ビールを10回以上購入したユーザのログで占められている。そのため、本実験では10回以上の購入者1,763人に絞って分析を行った。またデータの情報としては「ユーザID」「購入時期」「購入商品JANコード」を利用し、購入時期を週単位に変換を実施。該当ユーザの平均購入回数はおおよび、中央値を表1に示した。また、ユーザとビール購買数の分布を図4に示した。

次に、ビールのアイテムに対しては、商品JANコードより手動で8つのブランドに分類を行った。JANコードとは、商品毎に標準で13桁のIDが割り振られ、IDの番号から生産国や生産会社を推定することができる。そこで本研究では上位7桁の番号でブランドを割り振り、最終的には1,404の商品を、AからHの8つのブランドに分類。各ブランドの商品数の概算を表2に示した。なお

表 1 利用データ

データ利用期間	2013/1 - 2013/12
ビール品目 10 回以上購入者数	1,763 人
10 回以上購入者のログ数	77,926
ユーザ平均購入回数	44.2 回
ユーザ購入回数中央値	27 回
商品種類数	1,404
アイテム平均購入回数	55.5 回

図 4 ビールの購入頻度分布と購入者数 .

本稿では、分類方法とブランド名の詳細は省略する .

表 2 各ブランドと商品概算数

ブランド	商品数
A	約 250
B	約 200
C	約 180
D	約 30
E	約 200
F	約 200
G	約 200
H	約 100
総計	1,404

図 5 はブランド毎の購買数を、週の粒度でグラフ化したものであり、例えばブランド B では GW には一定のビール購買増加が見られることがわかる . また図 6 は、購買数の高いブランド A, B, C, E の 4 ブランドに対してブランド毎の購買変動 (a) および変動積算値 (b) を図示したものである . ここで購買変動とは、ブランド i の購買数 $S_i(t)$ に対して購買平均を $\mu_i = 1/T \sum_{t=1}^T S_i(t)$, 標準偏差を $\sigma_i = \sqrt{1/T \sum_{t=1}^T (S_i(t) - \mu_i)^2}$ とした時に、購買平均からの変動を、購買変動 $L_i(t)$ として、また、購買変動の積算値を $M_i(t)$ として下記のように与えた .

$$L_i(t) = (S_i(t) - \mu_i) / \sigma_i \quad (1)$$

$$M_i(t) = \sum_{t'=1}^t L_i(t') \quad (2)$$

ここで、本件では 1 年間のデータのため $T = 53$ である . 図 6(b) より、ブランド A, B, C では冬場の 11 月から春の 3 月までは平均的に購買数が抑えられ、初夏から秋に

図 5 ブランド毎の購買頻度を週単位で表示

図 6 主要 4 ブランドの購買変動 (a) および変動積算値 (b) を表示した .

かけての購買数が多くなっていることがわかる . ブランド A では 3 月より購買数を伸ばし、10 月まで購買数が高くなっている . ブランド B では 5 月の GW 時期に急激に購買数を伸ばし、11 月まで購買増を継続し、ブランド C では 3 月中旬より徐々に購買数を伸ばし、11 月まで購買増を継続していることがわかる . 一方、ブランド E では、3 月や 5 月に一時的に購買増があるものの、夏場に大きく購買数を伸ばすこともなく、かつ冬場での購買減の影響も低いかわがわかる . このように、ブランド毎に購買傾向の季節変動が観測されることがわかる .

3. NTF によるパターン抽出

一般的に購買履歴などの構造化されたデータはテンソルとして表現することが可能であり、これらのテンソル表現されたデータのパターン抽出手法として、近年 NTF

図7 NTF の概念図． $X \approx A \otimes B \otimes C$ となるように因子分解を行う．

が広く利用されている [Takeuchi 13b, 松林 14]．本研究では NTF を用いてパターン分析を行い (ユーザ $[I]$) $\times$ (時間 $[J]$) $\times$ (アイテム $[K]$) の 3 次テンソルデータ $X = [x_{ijk}] \in \mathbb{R}_+^{I \times J \times K}$ として因子分解を行った．本研究では時間情報の粒度に関しては最も定性評価に適した週 ($J = 53$) 単位を用いた．各パラメータの値は, $I = 1763$, $J = 53$, $K = 1404$ である．

本研究では一般的に用いられている NTF の CANDECOMP/PARAFAC decomposition (CP 分解) [Kolda 09, Cichocki 09] を行い, 3 次テンソル X を, ランク数を R とした 3 個の行列 $A = [a_{ir}] \in \mathbb{R}_+^{I \times R}$, $B = [b_{jr}] \in \mathbb{R}_+^{J \times R}$, $C = [c_{kr}] \in \mathbb{R}_+^{K \times R}$ の内積として分解する． A, B, C は各要素の分解行列であり, 分解行列の A, B, C を用いて行列内積を下記とした時,

$$\hat{X} = [\hat{x}_{ijk}] = A \otimes B \otimes C \quad (3)$$

として得られるテンソル $\hat{X}$ に対して,

$$D = \sum_i \sum_j \sum_k d(x_{ijk}, \hat{x}_{ijk}) \quad (4)$$

を最小化するように, A, B, C を求める．ここで, $d(x_{ijk}, \hat{x}_{ijk})$ は距離関数で, 通常この距離関数には KL ダイバージェンスやユークリッド距離が用いられる．一般的には 5 段階評価などの数値にばらつきが小さいデータにはユークリッド距離が用いられ, ログや信号処理などの数値の取りうる範囲が大きいデータには KL ダイバージェンスが用いられる．そのため本研究では KL ダイバージェンスを用いる．また, $\hat{X} = [\hat{x}_{ijk}] \in \mathbb{R}_+^{I \times J \times K}$ であり, $\hat{x}$ は下記として得られる．

$$\hat{x}_{ijk} = \sum_r a_{ir} b_{jr} c_{kr} \quad (5)$$

NTF とはこれら A, B, C の各要素の非負性を保ったまま, この D を最小化させる行列 A, B, C を求める手法である．図 7 に NTF 分解のイメージを示した．本研究では, ランク数 R に関して, $R = 5, 6, 8, 10, 20, 30, 50, 100$ とパラメータを振って分析をした結果, 時系列の特徴として定性的に評価を行い $R = 5$ を選択した．

4. 結果と考察

4.1 NTF 因子分解結果

図 8 は NTF による各因子行列の結果である．各行列の要素は正規化し, 各行列は以下の通りである．

$$\bar{A} = [\bar{a}_{ir}] = \sum_j \sum_k \frac{x_{ijk}}{\hat{x}_{ijk}} t_{ijk} \quad (6)$$

$$\bar{B} = [\bar{b}_{jr}] = \sum_i \sum_k \frac{x_{ijk}}{\hat{x}_{ijk}} t_{ijk} \quad (7)$$

$$\bar{C} = [\bar{c}_{kr}] = \sum_i \sum_j \frac{x_{ijk}}{\hat{x}_{ijk}} t_{ijk} \quad (8)$$

ここで $t_{ijk} = a_{ir} b_{jr} c_{kr}$ である．これらの正規化によって, 各行列要素に対し, 実際の相当レコード数が各行列に与えられ, 分析結果の解釈に有効となる．これらの結果から, $\bar{B}$ 行列を線グラフ化したものが図 9 であり, 一部のランクでは購買時系列パターンに特徴が見られる．

図 10 は各ランクの購買変動積算値を示し, 式 (2) と同様に, 平均値からの乖離を, 標準偏差で正規化した値による変動値を示している．図 9 および図 10 からわかることは, ランク 1 は他のランクによる時系列と比較しても冬場の購買数が低く, 特に 11 月から 2 月購買数が低い．またランク 2 は他と比較しても夏の購買数が大きく伸び, 8 月と 9 月の購買が好調である．一方で, 10 月以降は購買数が伸びていないことも理解できる．ランク 3 とランク 5 では共に春先か購買数が増加し, 共に 7 月と 8 月の夏場に購買数が伸びず, ランク 3 では 9 月に一度購買数が落ちるものの 11 月には再び購買数を上げていることもわかる．ランク 4 はランク 1 と似たような購買数の変動傾向が見られるものの, 5 月の GW 時期に購買数が一時的に増加し, 夏の間も一定の購買増が見られる傾向がある．年間変動の大きさは, 図 10 において線分の傾きが大きくなる時期に該当し, ランク 3 と 5 は乖離の振れ幅が小さく, 年間の変動が比較的小さいことがわかり, ランク 1, 2, 4 は年間の変動が比較的に大きいことがわかる．

図 11 はランク毎の購買数 $\bar{c}_{kr}$ を集計した値を棒グラフ化したものであり, さらにブランド毎に細分化した内訳を示している．また, 図 12 はブランド毎の購買数 $\bar{c}_{kr}$ を集計し, さらにランク毎とそのブランドの内訳比率を示している．両図から読み取れることは, 例えば, ブランド E, F ではランク 2 の比率が低いことから, ブランド戦略的に夏場や余暇時での購買をターゲットとした商品が少ないことが読み取れる．また, 各ブランドにおいて季節変動の比較的小さいランク 5 の購買数が高く, ランク 3 の購買数が低いことは, 全体的な傾向としては春先から夏場にかけての購買増加が見られるものの, ランク 3 に見られるような傾向のアイテムは希少であることがわかる．特に秋に購買数が再び増加するため, 秋にも

図8 NTFの分析結果．左より $\bar{A}$ 行列（ユーザ × ランク）， $\bar{B}$ 行列（時間 × ランク）， $\bar{C}$ 行列（アイテム × ランク）．各図共に，横軸がランクである．各行列値の大きさの最小値（ここでは 0）と最大値の間に，最右図に示すカラーマップを与えている．

購買数増加が見込める可能性もある．特に，ブランド D ではランク 3 の購買数が 0 で，独自に季節性のある商品がないことを示している可能性がある．

5. グラフ可視化による考察

本章では，各ランクでのアイテム分布と，各ブランドでのアイテムおよびランク分布の特徴を考察するため，グラフ可視化を用いた定性的評価を行った．

初めに，アイテムの購買特徴のパターン類似度をランク比率から算出し，商品類似度からグラフデータを作成した．アイテム k と k' の類似度を，コサイン類似度

$$Sim_{kk'} = \cos(\bar{c}_{kr}, \bar{c}_{k'r}) = \frac{\sum_r \bar{c}_{kr} \bar{c}_{k'r}}{\sqrt{\sum_r \bar{c}_{kr}^2} \sqrt{\sum_r \bar{c}_{k'r}^2}} \quad (9)$$

より求め，全アイテム間の類似度を算出．次に， $Sim_{kk'} \geq 0.92$ でカットオフを行い，グラフデータから孤立ノードを除外し，最終的にノード数 1,305，エッジ数 38,025 からなるグラフデータを抽出した．図 13 および図 14 はグラフデータを元にグラフ可視化[松林 07]を行った結果である．両図では $\bar{c}_{kr}$ の値に対してノードサイズを変え，かつ図 13 では，各ノードに対して $\bar{c}_{kr} / \sum_r \bar{c}_{kr} \geq 0.3$ となるランク r のみを色付けして描画させている．

次に，各ランク間の類似度も同様に，ランク r と r' に対してコサイン類似度， $\cos(\bar{c}_{kr}, \bar{c}_{k'r'})$ ，から算出し，表 3 にまとめた．この結果からは，ランク 2 とランク 3 の傾向をもつアイテムが比較的に多く，ランク 2 とランク 3 との共起度が高く，一方ランク 1 とランク 5 は共起度が低いことがわかる．

ランク分類可視化結果

図 13 では，各ランクへの分類結果より，ランク毎に似たような傾向を持つアイテムが抽出されていることがわかり，ランク間では少数のアイテムが複数のランクの特徴を持つ分布を持っている事が観測できる．ランク 2 の傾向が強いアイテムは夏場の売上が高く，比較的に固まって抽出されている傾向が見られる．ランク 3 の傾向

図9 $\bar{B}$ 行列（時間 × ランク）の各ランク r をグラフ化したもの．細線はベジエ曲線補完であり，各ランクの平均的な傾向を示している．

図10 $\bar{B}$ 行列（時間 × ランク）の各ランクの変動積算値を示す．

図 11 ランク毎の各ブランドの購買数

図 12 ブランド毎の購買数におけるランク比率

表 3 ランク間の共起度

	r=1	r=2	r=3	r=4	r=5
r=1	—	0.754	0.762	0.802	0.608
r=2	0.754	—	0.890	0.692	0.713
r=3	0.762	0.890	—	0.801	0.803
r=4	0.802	0.692	0.801	—	0.781
r=5	0.608	0.713	0.803	0.781	—

が強いアイテムは、ランク 5 やランク 1 の特徴を併せ持つノードも散見され、一般的に春先から秋にかけて定期的に売れている商品だと考えられるアイテムが見られる。ブランド分類可視化結果

図 13 では中心領域にアイテムが出現しなかった一方、図 14 では多くのアイテムが出現し、また購買数も平均的に高いものが多く存在する。図 13 ではアイテムの描画に対して、ランク比率による閾値を与えているため、一部のアイテムを描画していない。両図を同時に比較することにより、特に強い時系列パターンを持たず、通年で消費傾向が見られるアイテムが多く存在するということがわかる。ここで、図 14 では、特徴的なアイテムに関していくつか抽出し、図中に解釈を載せた。例えば、ランク 3 の比率が高かった領域のアイテムには、実際に秋をターゲットとした商品などが多く抽出されており、図 10 にてランク 3 のグラフが 10 月以降に再び購買数が増加していることも、理解することができる。ブランド A, B, C の主要銘柄でもランク 3 の比率が高い商品があるのは、該当商品は秋にも売れる傾向があることが考えられる。また、ランク 2 の比率が高いアイテムは、主要銘柄の他、輸入品や地ビールが多く観測された。これは、夏場などの旅行先など、普段とは異なる嗜好品を購入する傾向が見られている可能性がある。

以上の通り、NTF による因子分解を行い、さらに、各ランクへの比率を用いたグラフ可視化を行うことによって、各ランクへの分解結果と、ランク比率を用いたアイテムの定性的な観測を行うことが可能となり、両手法を組み合わせることが、ブランド選択分析を定性的な評価を行うのに有効であることが確認できた。

6. ま と め

本研究では、非負値テンソル因子分解 (NTF) を用いて、アイテム購買分析を行った。NTF ではデータ構造を 3 次元以上のテンソルとして表現することによって (ユーザ) $\times$ (時間) $\times$ (商品) といった高次の因子分解が可能になり、本研究では、ユーザ単位の購買傾向と時系列的特徴と個々の商品購買傾向を、合わせて因子分解を実施した。また、商品カテゴリをビール購買に絞り、商品のランク比率を用いたグラフ可視化を行うことによって、各ランクへの分解結果と、ランク比率を用いたアイテムの定性的な観測を実施し、アイテム購買の季節性の確認を行った。本研究では、NTF による高次因子分解分析と、グラフ可視化によるアイテムの俯瞰観察手法の両手法を組み合わせることによって、ブランド選択分析に対する一つの分析手法としての有用性の確認を行った。

しかしながら、購買行動における要因には様々な粒度が存在し、分析者が恣意的に分析の粒度を決めなければならない。また、NTF 等の因子分解手法は、元データに対して低ランク近似が可能であることを前提条件として

図 13 アイテムのランク類似度を元にグラフ可視化結果を行った結果 1 . 各ノードに対してランク r ごとに色分けを行い, 各ランクへの寄与度が高いノードのみを表示し, 購買数に比例してノードサイズで描画を行っている .

図 14 アイテムのランク類似度を元にグラフ可視化結果を行った結果 2 . 各ノードに対してブランドごとに色分けを行い, 購買数に比例したノードサイズで描画を行っている .

いるため、単純な NTF による因子分解手法では、全体的に高い傾向のパターンのみが抽出されてしまう。それ故、自動的にランク数 R を決定するノンパラメトリック手法や、特に購買分析などに特化した特徴量抽出のための定量的な指標を決定する必要性がある。しかしながら、因子分解技術においてこのような指標の決定方法や、そのアルゴリズムの確立は非常に難しい問題であるため、今後の研究課題として取り組んでゆく。また、ブランド選択問題としては、より大規模なデータによるアプローチや、ブランドスイッチングなどの分析も可能な階層的特徴抽出技術など、今後は、NTF の拡張手法を開発を目指してゆく。

◇ 参 考 文 献 ◇

- [Guadagni 83] Guadagni, Peter M., and John DC Little. "A logit model of brand choice calibrated on scanner data." *Marketing Science* 2.3 (1983): 203-238.
- [Winer 86] Winer, Russell S. "A reference price model of brand choice for frequently purchased products." *Journal of Consumer Research* (1986): 250-256.
- [hardie 93] Hardie, Bruce GS, Eric J. Johnson, and Peter S. Fader. "Modeling loss aversion and reference dependence effects on brand choice." *Marketing Science* 12.4 (1993): 378-394.
- [Revelt 98] Revelt, David, and Kenneth Train. "Mixed logit with repeated choices: households' choices of appliance efficiency level." *Review of Economics and Statistics* 80.4 (1998): 647-657.
- [Andrews 02] Andrews, Rick L., Andrew Ainslie, and Imran S. Currim. "An empirical comparison of logit choice models with discrete versus continuous representations of heterogeneity." *Journal of Marketing Research* 39.4 (2002): 479-487.
- [Allenby 98] Allenby, Greg M., and Peter E. Rossi. "Marketing models of consumer heterogeneity." *Journal of Econometrics* 89.1 (1998): 57-78.
- [Arora 98] Arora, Neeraj, Greg M. Allenby, and James L. Ginter. "A hierarchical Bayes model of primary and secondary demand." *Marketing Science* 17.1 (1998): 29-44.
- [Erdem 96] Erdem, Tülin. "A dynamic analysis of market structure based on panel data." *Marketing Science* 15.4 (1996): 359-378.
- [近藤 99] 近藤文代. "デイリー POS データにおける曜日変動および値下げ効果の抽出." *オペレーションズ・リサーチ: 経営の科学* 44.3 (1999): 154-163.
- [阿部 05] 阿部誠, 近藤文代. "マーケティングの科学-POS データの解析-" 予測と発見の科学シリーズ, 朝倉書店, (2005)
- [Lee 01] Lee, Daniel D., and H. Sebastian Seung. "Algorithms for non-negative matrix factorization." *Advances in Neural Information Processing Systems* (2001)
- [澤田 12] 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" *電子情報通信学会学会誌*, 95(9) (2012): 829-833
- [Takeuchi 13a] Takeuchi K., Ishiguro K., Kimura A., and Sawada H. "Non-negative Multiple Matrix Factorization." *Proceedings of 23rd International Joint Conference on Artificial Intelligence (IJ-CAI2013)*, pp.1713-1720, 2013.
- [幸島 13] 幸島 匡宏, 江崎 健司, 高屋 典子, 澤田 宏. "複数列の行列分解に関する一考察," 第 16 回情報論的学習理論ワークショップ (IBIS2013), 2013.
- [Shashua 05] Shashua A., and Tamir H. "Non-negative tensor factorization with applications to statistics and computer vision." *Proceedings of the 22nd International Conference on Machine Learning (ICML2005)*, ACM, 2005.
- [大石 06] 大石敦志. "日別 POS データによるビール・発泡酒の需要分析." *食品経済研究* 34 (2006): 64-71.
- [土井 14] 土井千章, 小西哲平, 中川智尋, 片桐雅二, 稲村浩, 太田賢. "店舗クラスタを用いた購買行動に関するライフスタイルの推定." *DEIM Forum 2014*, F3-5 (2014)
- [Takeuchi 13b] Takeuchi, Koh, et al. "Non-negative Multiple Tensor Factorization." *Data Mining (ICDM), 2013 IEEE 13th International Conference on*. IEEE (2013)
- [松林 14] 松林達史, 幸島匡宏, 林亜紀, 澤田宏. "非負値テンソル因子分解を用いたパターン抽出とその応用例." 第 11 回ネットワーク生態学研究会 (Neteco2014) (2014)
- [Kolda 09] Kolda, Tamara G., and Brett W. Bader. "Tensor decompositions and applications." *SIAM review* 51.3 (2009): 455-500.
- [Cichocki 09] Cichocki, Andrzej, et al. "Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation." John Wiley & Sons (2009)
- [松林 07] 松林達史, et al. "固有楕円ポテンシャルを利用したラベル付きグラフ可視化の座標計算." *情報処理学会研究報告. MPS, 数理モデル化と問題解決研究報告* 2007.128 (2007): 65-68.

〔担当委員: 鳥海 不二夫〕

2014 年 10 月 7 日 受理

——著 者 紹 介——

松林 達史

2000 年, 京大物理学部物理学科卒業. 2002 年 10 月より 2 年半, 理化学研究所非常勤研究員. 2005 年, 東京工業大学大学院理工学研究科地球惑星科学専攻博士課程修了. 同年, 日本電信電話株式会社入社. 以来, NTT コミュニケーション科学基礎研究所, NTT サービスエボリューション研究所にて, グラフ可視化および機械学習の研究開発に従事. 博士 (理学). 情報処理学会会員.

幸島 匡宏

2009 年, 東京工業大学工学部情報工学科卒業. 2012 年, 同大学院総合理工学研究科知能システム科学専攻修士課程修了. 同年, 日本電信電話株式会社入社. 以来, NTT サービスエボリューション研究所にて, 機械学習, データマイニングの研究に従事.

林 亜紀

2010 年, お茶の水女子大学理学部情報科学科卒業. 2012 年, 同大学院人間文化創成科学研究科博士前期課程修了. 同年, 日本電信電話株式会社入社. 以来, NTT サービスエボリューション研究所にて, 可視化の研究開発に従事. お茶の水女子大学大学院人間文化創成科学研究科理学専攻博士後期課程在学中. 日本データベース学会会員.

澤田 宏

1991 年, 京都大学工学部情報科卒業. 1993 年, 同大学院工学研究科修士課程修了. 同年, 日本電信電話株式会社 (NTT) 入社. 以来, 同社コミュニケーション科学基礎研究所にて, VLSI 向け CAD および計算機アーキテクチャの研究に従事. 2000 年より, 信号処理, 特にブラインド音源分離の研究に従事. 2009 年より, 知能創成環境研究グループリーダー. 2013 年より, サービスエボリューション研究所ヒューマンアナリシスプロジェクトグループリーダー, 現在に至る.

2001 年, 京都大学博士 (情報学). 電子情報通信学会, 日本音響学会, IEEE 各会員.