

Abstract

近年データ分析の分野において非負値行列因子分解 (NMF: Non-negative Matrix Factorization) が高い注目を集めている。NMF は、データを行列表現することによって、データの持っている潜在的パターン抽出とその要因分析を行うことが可能な技術として、幅広い分野で利用されている。NMF の高次拡張技術である非負値テンソル因子分解 (NTF: Non-negative Tensor Factorization) は、より高次のデータを扱うことのできる技術であり、様々なデータに対して要因をより詳細に、かつ多角的に分析することが可能であり、近年更に注目を集めている。例えば、マーケティングサイエンスにおける購買ログ分析では、ユーザの商品嗜好分析だけでなく、嗜好の移り変わりや、販促時期や購買店舗の分析など、より詳細かつ多角的な要因分析が必要とされている。三次以上のテンソルデータを分析可能な NTF では、(ユーザ) × (商品) × (購買店舗) を同時に因子分解することによって、「子供のいる主婦層はスーパーマーケットで食玩やお菓子をよく購入する」などの、複数要因の同時分析が可能となる。一方、分析因子要素を増やすことは同時に計算量の増加にもつながる。実世界のデータ分析においては疎 (スパース) なデータを扱うことも多く、アルゴリズムや実装上の工夫により計算を効率化させることが必要不可欠である。本稿では、スパースデータの非負値テンソル因子分解技術に関する定式化及び実データを用いた分析技術を合わせて紹介する。

キーワード: 非負値テンソル因子分解, スパースデータ, 高速化

1. はじめに

近年、携帯電話やセンサデバイスの技術向上に伴い、取得可能なデータの容量と多様性が増加してきている。

目次

- [I] 複合データ分析技術とその発展 (6月号)
- [II・完] テンソルデータの因子分解技術と実応用例 (7月号)

特にビッグデータの利活用という観点からは、取得データの持つ多くの属性を考慮した分析技術が重要視され、高次の情報を扱う機械学習技術の注目度が非常に高い。とりわけ高次情報の表現手法である「テンソル」を利用した非負値テンソル因子分解 (NTF: Non-negative Tensor Factorization) は、近年高い注目を集め、かつ様々な応用技術が提唱されている。本稿では、この NTF という技術とその発展に関して、基本的な解説とともに、利用シーンや利点などを実データ分析例を用いて紹介を行う。

2. 非負値テンソル因子分解 (NTF)

連載第1回から示してきたように、一般的に様々なデータは行列として表現することができる。例えば図1に見られるように、購買ログのユーザと商品情報を考慮したデータは行列として表現できる。更に、購入場所 (店舗) を考慮すると、データは三次のテンソルとして表現することができ、同じ書籍やパソコンの購買に関しても、「CさんはECサイトをよく利用するので、ECサ

松林達史 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
E-mail matsubayashi.tatsushi@lab.ntt.co.jp
幸島匡宏 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
E-mail kohjima.masahiro@lab.ntt.co.jp
澤田 宏 正員 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
E-mail sawada.hiroshi@lab.ntt.co.jp
Tatsushi MATSUBAYASHI, Masahiro KOHJIMA, Nonmembers, and Hiroshi SAWADA, Member (NTT Service Evolution Laboratories, NIPPON TELEGRAPH AND TELEPHONE CORPORATION, Yokosuka-shi, 239-0847 Japan).
電子情報通信学会誌 Vol.99 No.7 pp.691-698 2016年7月
©電子情報通信学会 2016

年月日	ユーザ	年代性別	商品	場所	購入数
2013年2月5日	Aさん	30代男性	ノートPC	家電量販店	2
2013年2月5日	Bさん	40代女性	参考書	本屋	3
2013年2月8日	Cさん	20代男性	ピザ	スーパー	2
2013年2月11日	Bさん	30代男性	ピザ	スーパー	1
2013年2月14日	Cさん	40代女性	参考書	ECサイト	1
2013年2月14日	Cさん	40代女性	ノートPC	ECサイト	1
⋮	⋮	⋮	⋮	⋮	⋮

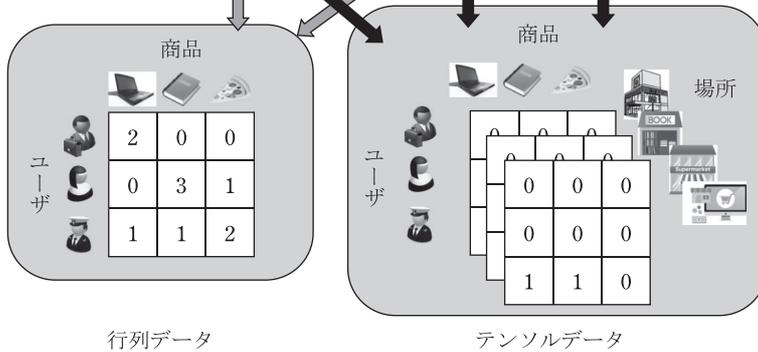


図1 データ属性の増加とテンソル化

イトならではの商品推薦をした方がよい」といった分析が可能となる。

2.1 NTF の定式化

一般に、 N 個のモードを持つテンソルデータを、 N 次のテンソルデータという。例えば、非負値の実数値を持つ三次のテンソルデータは $X = [x_{ijk}] \in \mathbb{R}_+^{I \times J \times K}$ と表すことができる。ここで I, J, K は各モードの要素数で、 \mathbb{R}_+ は非負の実数値を表す。

代表的なテンソル因子分解手法の一つである CP 分解では、 X を基底数 R 個の因子に分解を行う。このとき、各因子と要素の関係は因子行列として表現ことができ、各モードの因子行列 A, B, C はそれぞれ $A = [a_{ir}] \in \mathbb{R}_+^{I \times R}$, $B = [b_{jr}] \in \mathbb{R}_+^{J \times R}$, $C = [c_{kr}] \in \mathbb{R}_+^{K \times R}$ として表せる。因子行列 A, B, C の積を $\hat{X} = [\hat{x}_{ijk}] \in \mathbb{R}_+^{I \times J \times K}$ として以下のように表記できる。

$$\hat{x}_{ijk} = \sum_{r=1}^R a_{ir} b_{jr} c_{kr} \quad (1)$$

この \hat{X} を、 X との誤差が小さくなるように因子行列を求める。非負値行列因子分解 (NMF: Non-negative Matrix Factorization) は行列積で \hat{X} を表したのに対し、NTF では基底ごとの外積の和を用いる。NMF と NTF の因子分解イメージを図2に示した。

通常 \hat{X} と X との差分は誤差関数を用いて定義され、この誤差関数 $\mathcal{D}(\cdot)$ を、一般的には二乗誤差や一般化

$$I \times J \times X \approx I \times A \times R \times B^T \times J$$

$$\hat{x}_{ij} = \sum_{r=1}^R a_{ir} b_{jr}$$

(a) NMF: 非負値行列因子分解

$$I \times J \times K \times X \approx I \times A \times R \times B \times J \times C \times K$$

$$\hat{x}_{ijk} = \sum_{r=1}^R a_{ir} b_{jr} c_{kr}$$

(b) NTF: 非負値テンソル因子分解

図2 NMF と NTF の分解のイメージ図

KL ダイバージェンスで与える。例えば、一般化 KL ダイバージェンスを誤差関数としたときは、以下である。

$$\mathcal{D}(X \parallel \hat{X}) = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left(x_{ijk} \log \frac{x_{ijk}}{\hat{x}_{ijk}} - x_{ijk} + \hat{x}_{ijk} \right) \quad (2)$$

最終的には因子行列の非負性を保ちつつ、 $\mathcal{D}(\cdot)$ を最小化させると更新式は以下のように得られる。

$$\text{NTF の更新式} \left\{ \begin{array}{l} a_{ir} \leftarrow a_{ir} \frac{\sum_{j=1}^J \sum_{k=1}^K \frac{x_{ijk}}{\hat{x}_{ijk}} b_{jr} c_{kr}}{\sum_{j=1}^J \sum_{k=1}^K b_{jr} c_{kr}} \\ b_{jr} \leftarrow b_{jr} \frac{\sum_{i=1}^I \sum_{k=1}^K \frac{x_{ijk}}{\hat{x}_{ijk}} a_{ir} c_{kr}}{\sum_{i=1}^I \sum_{k=1}^K a_{ir} c_{kr}} \\ c_{kr} \leftarrow c_{kr} \frac{\sum_{i=1}^I \sum_{j=1}^J \frac{x_{ijk}}{\hat{x}_{ijk}} a_{ir} b_{jr}}{\sum_{i=1}^I \sum_{j=1}^J a_{ir} b_{jr}} \end{array} \right. \quad (3)$$

なお、誤差関数と更新式の導出に関しては、文献(1)と連載第1回の記事を参照して頂きたい。

2.2 実データへの復元化手法

因子分解技術は、データを R 個の基底にパターンに分類することが可能である一方、低ランク性を仮定しているために \mathbf{X} と $\hat{\mathbf{X}}$ には誤差が生じる。実データの分析においては、定性的な解釈を行うために、元データとの整合性を考慮する必要性が生じることがある。そこで本節では、因子分解した後の因子行列に対して、この整合性を取るという目的で「誤差を戻す」という復元化手法の紹介を行う。

例えば因子行列 \mathbf{A} の要素 a_{ir} は、 $t_{ijkr} = a_{ir} b_{jr} c_{kr}$ を用いて以下のようにして復元化が可能である。

$$\bar{a}_{ir} = \sum_{j=1}^J \sum_{k=1}^K \frac{x_{ijk}}{\hat{x}_{ijk}} t_{ijkr} \quad (4)$$

\bar{a}_{ir} は、 R に対する和を取ると要素 i に対する元データの総レコード数が復元でき、下記の関係が成り立つ。

$$i \text{ のレコード総和} = \sum_{r=1}^R \bar{a}_{ir} = \sum_{j=1}^J \sum_{k=1}^K x_{ijk} \quad (5)$$

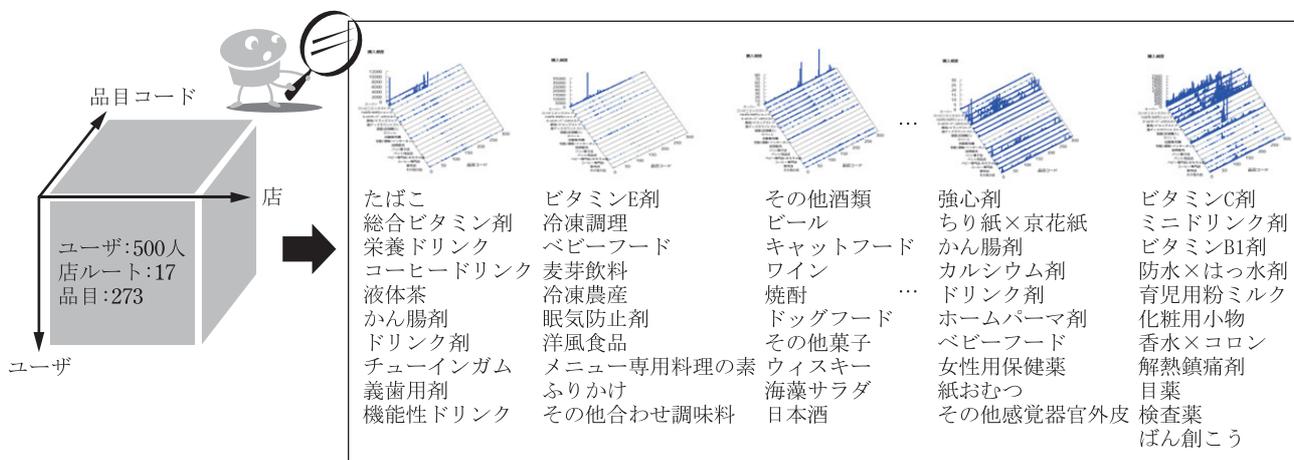


図3 ユーザと購買品と購買場所を考慮した、NTFによる分析例⁽²⁾ 基底数 $R=10$ として10種類の購買パターンを抽出した。

また \bar{a}_{ir} は、 I と R に対する和を取ると全レコードの総和 $(= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K x_{ijk})$ となり、

$$\sum_{i=1}^I \sum_{r=1}^R \bar{a}_{ir} = \sum_{j=1}^J \sum_{r=1}^R \bar{b}_{jr} = \sum_{k=1}^K \sum_{r=1}^R \bar{c}_{kr} \quad (6)$$

が成り立つ。 \bar{b}_{jr} と \bar{c}_{kr} も同様に定義することができる。

また、因子行列の積も同様に定義することができ、例えば、 \bar{bc}_{jkr} は以下で与えることができる。

$$\bar{bc}_{jkr} = \sum_{i=1}^I \frac{x_{ijk}}{\hat{x}_{ijk}} t_{ijkr} \quad (7)$$

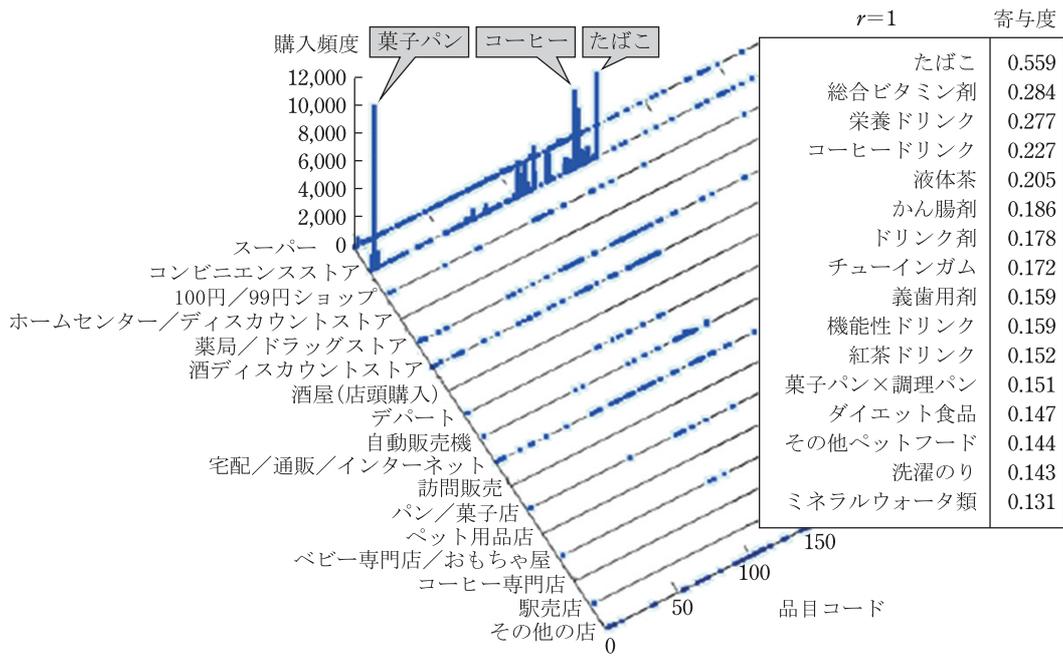
\bar{bc}_{jkr} も同様に、 R に対する和を取ると、要素 (j, k) に対する元データの総レコード数が復元でき、下記の関係が成り立つ。

$$(j, k) \text{ のレコード総和} = \sum_{r=1}^R \bar{bc}_{jkr} = \sum_{i=1}^I x_{ijk} \quad (8)$$

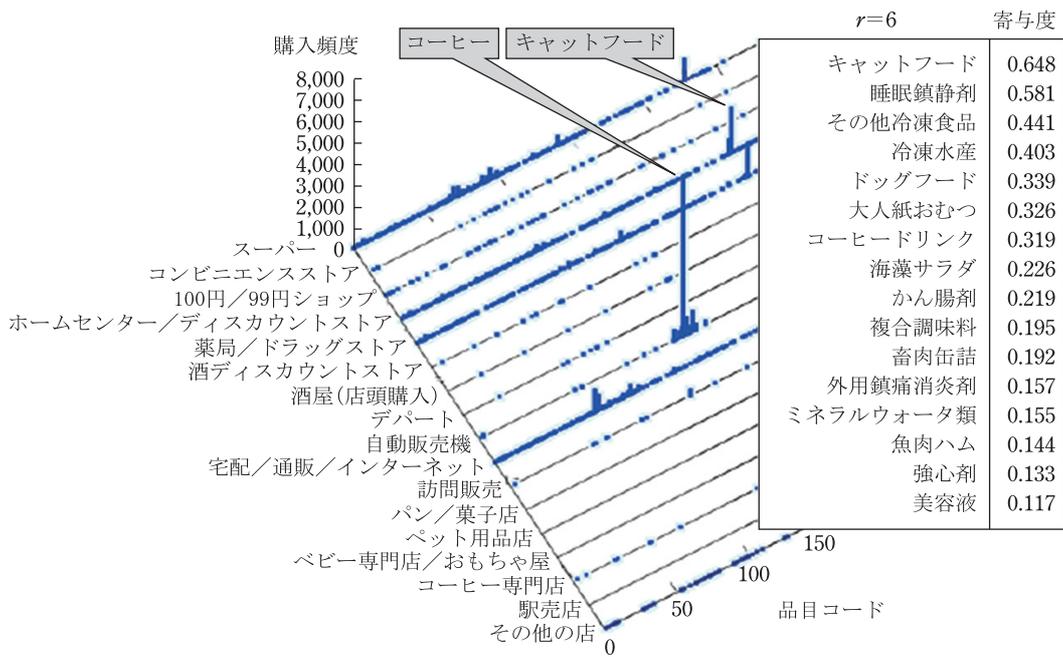
同様に、 J, K, R に対する和を取ると全レコードの総和になる。他の因子行列の積も同様に定義することができ、復元化された因子行列及び因子行列の積は、元データの数値に対応した値を取るため実際の値として比較しやすくなり、実データの分析において有用である。以下、実分析例と併せて紹介を行う。

2.3 NTF を用いた購買傾向分析

図3及び図4は、(株)インテージのスクランパネルデータ「SCI」を用いて、購買数の多い上位500ユーザーに絞り、 $R=10$ として10種類の購買パターンを抽出した結果である。「ユーザー情報 \mathbf{A} 、アイテム品目 \mathbf{B} 、購入場所 \mathbf{C} 」の三次テンソル情報を用いて分析を行った例⁽²⁾



(a) 基底1 ($r=1$)



(b) 基底6 ($r=6$)

図4 復元化された、基底1 ($r=1$) と、基底6 ($r=6$) の \bar{b}_{jkr} 行列の三次元プロット 各図の右表は、各アイテム品目の寄与度 \hat{b}_{jr} を大きい値順で並べた、上位16品目を示している。

であり、特徴的な複数の購買・併売パターンが現れている。図3は分析イメージで、右側の三次元バーチャートは、アイテム品目と購入場所に関しての復元化された \bar{b}_{jkr} 行列を示している。10個の基底のうち、特に「コーヒー」の購買傾向が高い二つの基底 ($r=1, 6$) を図4に示した。

図4に示されている「寄与度」は、各アイテム品目ごとの、各基底に対して正規化した値であり、 $\hat{b}_{jr} =$

$\bar{b}_{jr} / \sum_{r=1}^R \bar{b}_{jr}$ で与える。例えば、たばこの商品購買数の55.9%の購買は基底1の購買パターンに分類されていることを示している。基底1から、コンビニエンスストアにて、たばこを購入する層が併せてコーヒーや菓子パンを購入する傾向が確認できた。また基底2では、自動販売機でコーヒーを購入する層はキャットフードを購入する層に多く見られることが分かった。グラフの縦軸は

\overline{bc}_{jkr} で復元化した値を用いたことによって、実際の購買数が反映されている。

NTF を用いることにより、コーヒーを購入する層にも様々なパターンが存在し、テンソルを用いた高次の分析によって、より詳細なターゲット層を浮き彫りにすることが可能になる。また、本稿のような正規化手法を用いることによって、分類結果の定性評価時の解釈性が向上するという利点がある。

3. スパース非負値テンソル因子分解 (S-NTF) の定式化

ここまで NTF の定式化と、有用性に関して解説してきた。一方で、テンソルによる高次化により、データのスパース性の増加による計算時間の増加という問題点が浮上する。例えば、 $I \times J$ の大きさを持つ行列を、NMF を用いて基底数 R に分解を行う際、計算のオーダーは $O(IJR)$ になる。属性情報を追加し、 $I \times J \times K$ の大きさ

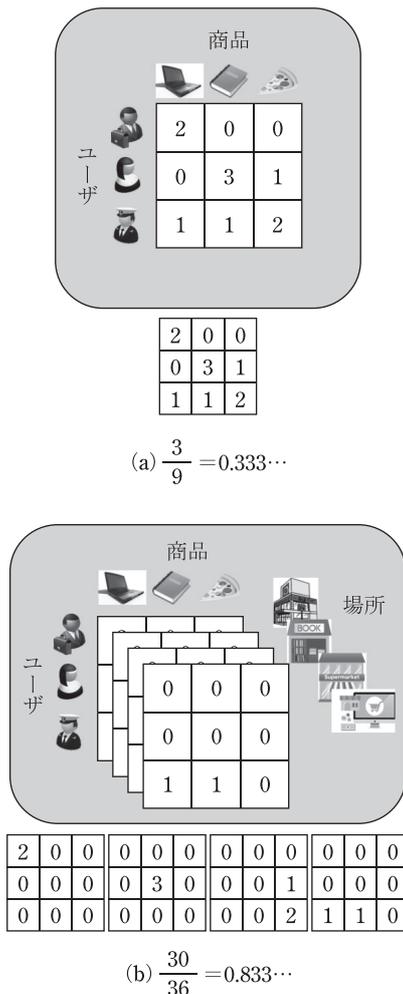


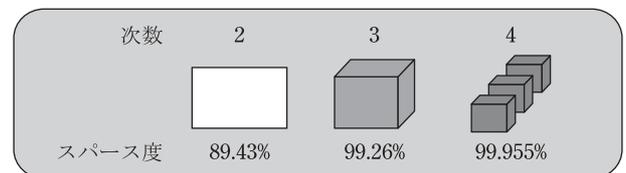
図5 図1の各データを展開したときのスパース度 元は同じデータでも行列のときは33.3%程度のスパース度が、三次テンソルでは83.3%と増加する。

を持つテンソルを、NTF を用いて基底数 R に分解を行う際には、計算のオーダーは $O(IJKR)$ になり、次数を上げると、追加したモードの次元数倍だけ計算量が増加する（この場合 K 倍になる）。

一方、購買ログなどの一般的なログでは、時間情報の粒度や、アイテムの総和数によって、非常にスパースなデータになる。例えば図5のように、元データは同じでも、分析の次数を増やすと、行列のときは33.3%程度のスパース度が、三次テンソルでは83.3%と増加することが分かる。図6は2.3の分析で用いた実データの例であるが、次数を増やすとスパース度が増加し、三次以上のテンソルデータは99%以上のデータ要素が“0”になっていることが分かる。したがって、“0”であることの計算を効率化させることが、処理の高速化へとつながる。

ここで本稿では、スパース非負値テンソル因子分解 (S-NTF: Sparse Non-negative Tensor Factorization) と呼ばれる、スパース度の高いデータに適した NTF の拡張技術を紹介する。まず始めに、「ユーザ (i) が、何を (j)、どこで (k) 購入したか」というデータ $(i, j, k) \in \mathbf{L}$ を考える。ログの集合を \mathbf{L} とし、ログ数 L ($=|\mathbf{L}|$) としたとき、 $(i, j, k) \in \mathbf{L}$ のときは $x_{ijk} = 0$ となるため、式(3)に代入すると、具体的な更新式は以下のように得られる、

$$\text{S-NTF の更新式} \left\{ \begin{array}{l} a_{ir} \leftarrow a_{ir} \frac{\sum_{(j,k) \in \mathbf{L}_i} x_{ijk} b_{jr} c_{kr}}{\sum_{j=1}^J \sum_{k=1}^K b_{jr} c_{kr}} \\ b_{jr} \leftarrow b_{jr} \frac{\sum_{(i,k) \in \mathbf{L}_j} x_{ijk} a_{ir} c_{kr}}{\sum_{i=1}^I \sum_{k=1}^K a_{ir} c_{kr}} \\ c_{kr} \leftarrow c_{kr} \frac{\sum_{(i,j) \in \mathbf{L}_k} x_{ijk} a_{ir} b_{jr}}{\sum_{i=1}^I \sum_{j=1}^J a_{ir} b_{jr}} \end{array} \right. \quad (9)$$



次数	組合せの数	1回以上の購買
2	「ユーザ」×「商品」 =1,422,280通り	150,316個
3	「場所」×「ユーザ」×「商品」 =24,178,760通り	179,324個
4	「商品」×「ユーザ」×「場所」×「時間」 =580,290,240通り	260,326個

図6 実データにおける、テンソルの次数とデータのスパース度

ここで L_i とは、 i に関する (j, k) のログの集合で、 L_j 、 L_k も同様に、 j の (i, k) に関するログと、 k の (i, j) に関するログの集合を示し、以下を満たす。

$$\sum_{i=1}^I |L_i| = \sum_{j=1}^J |L_j| = \sum_{k=1}^K |L_k| = L. \quad (10)$$

したがって、スパーステンソル因子分解における因子行列の更新には、 $O(LR)$ の計算量に抑えることができる。ここで図7は、C言語で実装したS-NTFのコードで、計算時間が L に比例していることが分かる。我々の計算コードで、Xeon E5-2670 2.6 GHz のCPUで、シングルスレッドによる処理で、400万ログのデータが1,000回の更新反復計算でも30分程度で終了できる。

S-NTFの実装により、2.3で行われた分析も100倍以上高速化され、2日ほど掛かっていた計算も1時間以内で分析が可能である。また、例えば時間要素を追加し、「ユーザ (i) が、何を (j)、どこで (k)、いつ (l) 購入したか」という分析をした場合を考える。時間窓を週ごとに分割し、時間変動を考慮すると1年間分のデータ分析で計算時間が53倍に増え、前述した分析も100日以上要する一方、S-NTFでは計算時間はほとんど増加せず、同様に1時間以内で分析が可能である。実際のコードは、大規模データ処理における、メモリのランダムアクセスによる実行速度の低下を抑えるなどの工夫をしているが、詳細は文献(3)を参照して頂きたい。

4. スパース非負値複合テンソル因子分解 (S-NMTF)

連載第1回では、複数のデータを同時に分析する非負値複合行列分解 (NMMF: Non-negative Multiple Matrix Factorization) についての技術紹介を行った。NMMFを用いることにより、複数のデータを利用した

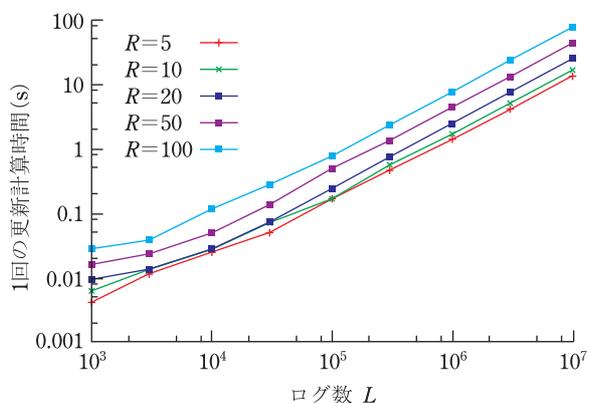


図7 S-NTFの実計算時間 横軸は分析ログ数、縦軸は1回の反復計算に要する時間。

より効率的な分析が可能になる。本章では更に、複数のテンソルデータを利用した分析技術、非負値複合テンソル因子分解 (S-NMTF: Sparse Non-negative Multiple Tensor Factorization) (4), (5)の適用例として、訪日外国人観光客を対象としたトライアルにおいて実際に行った分析結果と、プッシュ配信施策の紹介を行う。

4.1 訪日外国人観光客トライアルのクラスタリング分析

訪日外国人観光客トライアル (5), (6)とは、福岡市及び周辺観光地を訪れる外国人観光客を主な対象とするスマートフォンアプリケーション (観光アプリ) を配布し、無料Wi-Fiに簡単に接続できる機能や、ユーザの行動や状況に合わせて観光情報・割引クーポンなどのサービスを提供し、あらかじめ承諾を得たユーザから属性情報と位置情報、アプリの操作ログなどを収集した施策である。

収集したデータにS-NMTFを適用することによって、外国人観光客の回遊行動パターンの抽出や、エリアごとの観光客数の推定などの分析を実施した。本施策では五つのテンソルデータ及び行列データを利用し、それぞれ下記のとおりである。

- X_1 : {ユーザ, 空間メッシュ, 時間帯},
- X_2 : {ユーザ, 性別},
- X_3 : {ユーザ, 年代},
- X_4 : {ユーザ, 居住地域},
- X_5 : {ユーザ, メッシュ間移動歴}.

ここで分析ユーザ数は、属性が明確で位置情報取得が可能な240名に絞り、時間帯は1時間の時間帯で与え、空間は500m四方のメッシュで与えた。居住地域は「香港, 韓国, 日本, 台湾, 中国, アメリカ, その他」の7地域で与え、分析における基底数は予備実験の結果から $R=10$ としている。その他の詳細内容は文献(5)を参照して頂きたい。

クラスタリング結果の大きな傾向として、外国人観光客は福岡市市街地を中心に滞在するクラスタ群と、大分や熊本、長崎などの九州北部の都市を回遊するクラスタ群の二つに分かれることが分かった。図8は、各クラスタによる訪問頻度を地図上に表示したものである。

- 福岡市街地クラスタ

クラスタ0は佐賀県鳥栖にあるアウトレットモールにまで足を伸ばし、買い物をメインとした香港からの観光客が抽出された。またクラスタ4では、門司港と飯塚を中心とした、外国人観光客固有の観光地が抽出された。このクラスタでは台湾からの観光客が多く、台湾のテレビ番組で門司港の特集が組まれた影響とも推測できる。

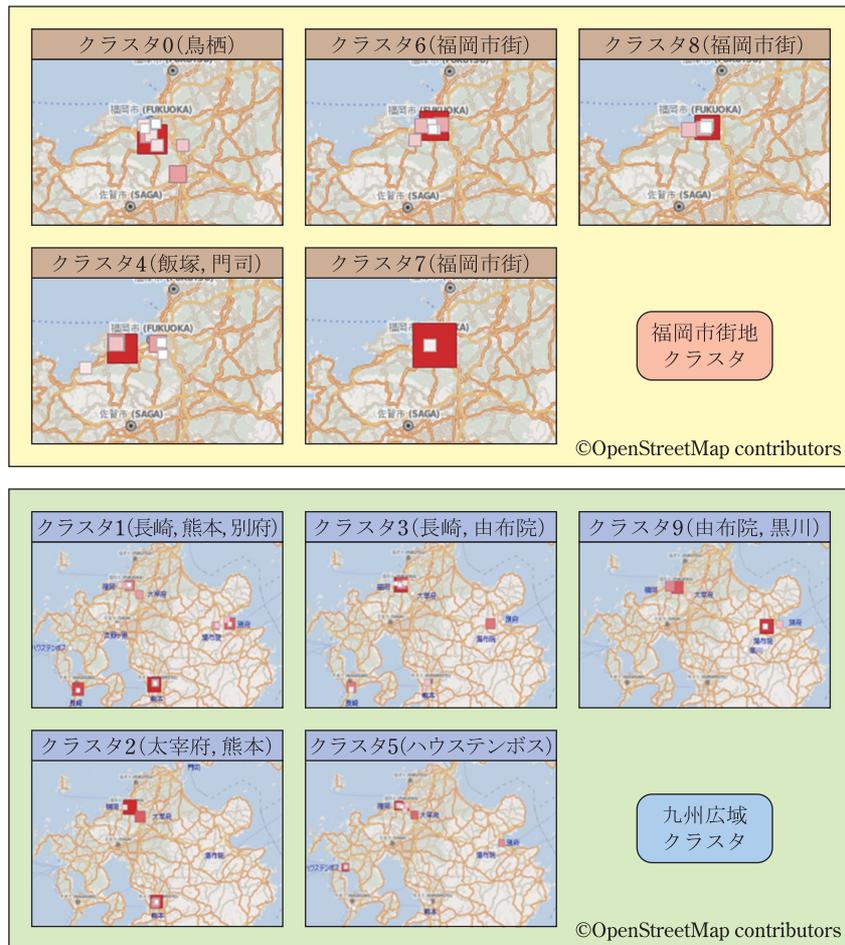


図8 各クラスタによる訪問頻度 分析の結果、クラスタリング結果の大きな傾向として、外国人観光客は福岡市市街地を中心に滞在するクラスタ群と、大分や熊本、長崎などの九州北部の都市を回遊するクラスタ群の二つに分かれることが分かった。

その他、福岡市街を中心とした外国人観光客のクラスタが抽出されている。

・ 九州広域クラスタ

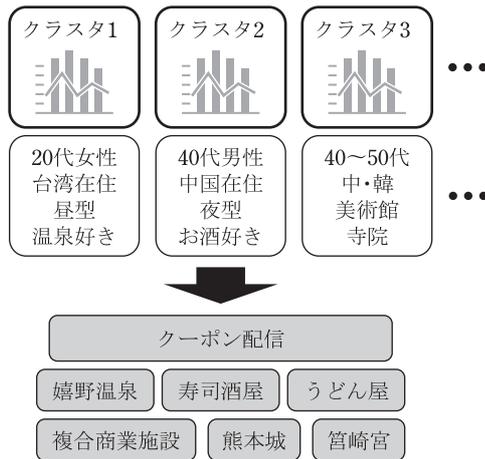
クラスタ1は、長崎、熊本、別府と九州北部の主要な観光地を大きく周遊するクラスタであり、韓国と香港からの観光客が多くを占め、ツアー旅行者の可能性が高い。クラスタ9は、大分県由布院及び熊本県黒川といった、有名な温泉街を回るクラスタが抽出された。このクラスタでは、台湾から来日した若い観光客を強く引きつけていることが分かった。その他、長崎のテーマパークや、太宰府などを中心とした観光客クラスタも抽出されている。

このように本施策では、有名な観光地が抽出されたほかにも、日本人には余りなじみがないが一部の外国人観光客には人気である観光地も発見された。また、滞留点を用いたユーザの位置情報履歴と同時に、ユーザ詳細な属性情報を考慮することによって、外国人観光客の中で

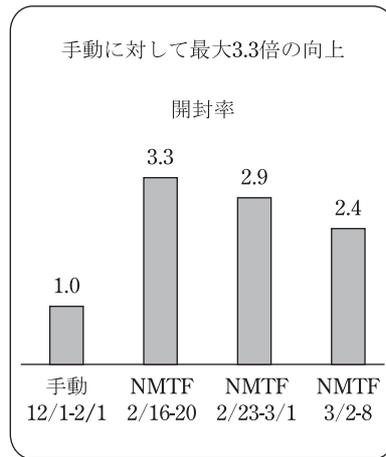
も、母国情報のみならず年代によって訪問する観光地に異なる潜在パターンが現れることが確認された。

4.2 ユーザクラスタリングを用いたプッシュ配信

本トライアルでは更に、ユーザクラスタリングの分類結果から、博多付近に訪れた外国人観光客に向けて観光情報・割引クーポンの内容を含んだ、スマートフォン向けプッシュ配信を実施した。図9は、事前分析による属性と配信内容の関係と、その配信施策のプッシュ配信通知の開封率である。前節のクラスタリング分析の結果を基に、各クラスタの特徴的なユーザ属性の組合せを抽出し、訪問位置情報と付近のPOI (Point of Interest) からユーザの興味カテゴリーの分析を行い、情報の配信内容をあらかじめ複数用意した。ユーザである外国人観光客が福岡を訪問した際に、該当するクラスタに属したユーザには、対応する情報を配信した。その結果、従来の手動による配信と比較して3回の配信実験において3倍ほどの開封率向上を達成した。



(a) 事前分析による属性と配信内容の分類の例



(b) プッシュ配信情報の開封率の結果

図9 クーポン配信内容と開封率

5. おわりに

2回にわたる連載では「複合データ分析技術とNTF」という題目で、非負値因子分解技術⁽¹⁾の様々な拡張技術紹介と、我々の実施したデータ分析結果や、外国人観光客に向けたプッシュ配信施策による分析の利活用方法を紹介してきた。非負値因子分解技術は、手法の理解と実装が比較的容易であり、かつ様々な応用手法が考案され続けている。更に実サービスにつながる分析手法としても利用が進んでおり、今後も長く利用され続ける技術になると、我々も考えている。

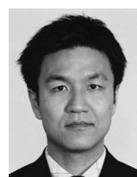
連載第1回では、複数のデータを効果的に分析に用いる、NMMFを中心とした分析技術の紹介を行った。今回は、行列に対してより高次の情報を扱うことが可能なテンソルを用いた、NTFの技術を中心に紹介を行った。NTFに関するより詳細な説明に興味を持つ方は、文献(7)と文献(8)を参照されたい。前者はNTFのサーベイ論文であり、NTF技術の理解に集中するのに適している。後者はNMFとNTF全般を扱った書籍で、かつ脳科学などへの応用例などが豊富で、因子分解技術の参考書として適している。

文 献

- (1) 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,” 信学誌, vol. 95, no. 9, pp. 829-833, Sept. 2012.
- (2) 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,” 11 回ネットワーク生態学シンポジウム, Sept. 2014.
- (3) 松林達史, 澤田 宏, “非負値スパーステンソル因子分解における Multiplicative 更新手法の高速化,” ハイパフォーマンスコンピューティングと計算科学シンポジウム論文集, pp. 81-81, 2015.

- (4) K. Takeuchi, R. Tomioka, K. Ishiguro, A. Kimura, and H. Sawada, “Non-negative multiple tensor factorization,” IEEE 13th International Conference Data Mining (ICDM), pp. 1199-1204, 2013.
- (5) 熊谷雄介, 今井良太, 松林達史, 佐藤吉秀, 堀岡 力, “非負値複合テンソル因子分解を用いた訪日外国人観光客の回遊行動分析,” 信学技報, IBISML 2015-3, pp. 15-19, June 2015.
- (6) 野口賢一, 佐藤吉秀, 塩原寿子, “高度高性能ビッグデータ活用技術とトライアル検証,” NTT 技術ジャーナル, vol. 27, no. 12, pp. 34-38, Dec. 2015.
- (7) T.G. Kolda and W.B. Brett, “Tensor decompositions and applications,” SIAM review, vol. 51, no. 3, pp. 455-500, 2009.
- (8) A. Cichocki, R. Zdunek, A.H. Phan, and S.I. Amari, Nonnegative matrix and tensor factorizations : applications to exploratory multi-way data analysis and blind source separation, John Wiley & Sons, 2009.

(平成 27 年 12 月 15 日受付 平成 28 年 1 月 6 日最終受付)



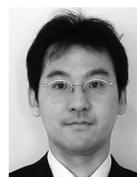
まつばやし たつし
松林 達史

平 12 京大・理・物理卒。平 14 から 2 年半、理研・非常勤研究員。平 17 東工大大学院理工学研究科博士課程了。同年日本電信電話株式会社入社。以来、データ分析技術の研究開発に従事。現在、NTT サービスエボリューション研究所主任研究員。博士(理学)。情報処理学会会員。



こうじま まさひろ
幸島 匡宏

平 21 東工大・工・情報卒。平 24 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了。同年、日本電信電話株式会社入社。以来、同社サービスエボリューション研究所にて、機械学習の研究に従事。



さわだ ひろし
澤田 宏 (正員)

平 3 京大・工・情報卒。平 5 同大学院修士課程了。同年、日本電信電話株式会社入社。以来、VLSI 向け CAD 及びアーキテクチャ、信号処理、機械学習の研究に従事。現在、同社サービスエボリューション研究所主幹研究員。平 13 京大博士(情報学)。日本音響学会、IEEE 各会員。