

複合データ分析技術と NTF [I] ——複合データ分析技術とその発展——

Multiple Data Analysis and Non-negative Matrix/Tensor Factorization [I] :
Multiple Data Analysis and Its Advances

幸島匡宏 松林達史 澤田 宏

Abstract

スマートフォンの普及やセンサデバイスの発達により、多岐にわたるデータの収集・蓄積が可能になった。そのため、これまでは考えられてこなかったデータの組合せを分析することで、新たな知見を発見できる可能性が高まっている。複合データ分析技術は、複数のデータを組み合わせることで、データ中に潜むパターン間の関係性を自動抽出することができる技術であり、近年のデータ分析の現場において広く利用されている。本稿では、画像、文書、音声など様々なデータ分析において有効性が示されている非負値行列因子分解 (NMF: Non-negative Matrix Factorization) を基礎とし拡張させた、複合データ分析技術の近年の発展について解説する。NMF の基礎の説明から複合データ分析の発展的課題まで幅広く紹介する。

キーワード：複合データ分析，非負値行列分解，非負値テンソル分解

1. はじめに

近年のデータ分析において解析対象となるデータの多くは、非負の値を要素に持つ行列、若しくは3軸以上の高次の配列 (テンソル) として表現することができる。例えば図1に示すように、文書集合は行と列がそれぞれ文書と単語に対応する、各文書中における単語の出現回数を表す行列で表現できる。誰がどの商品をどこで何個

購入したかを表す購買履歴も、行、列、奥行き、要素がそれぞれユーザ、店舗、商品、購入数を表すテンソルとして表現可能である。

本連載で紹介する複合データ分析技術の基礎となる手法、非負値行列分解 (NMF: Non-negative Matrix Factorization)^{(1), (2)}、非負値テンソル分解 (NTF: Non-negative Tensor Factorization)⁽³⁾は、行列若しくはテンソルで表現されるデータから、データ中に潜むパターンを自動で抽出する手法である。NMF 及び NTF の適用により、入力となる行列またはテンソルは、行列の積の形へ分解され、得られた分解結果がデータ中のパターンに対応する。例えば、購買データへ適用することでチョコレート好きやコーヒー好きなどの潜在的な購買のパターンを抽出し、各ユーザがどの購買パターンにどの程度基づいて購買するかを把握することが可能である。

複合データ分析技術は、NMF や NTF に対して単一のデータではなく、複数の行列及びテンソルデータを同時に分析することが可能となるように発展させた技術である。例えばユーザのファッションに関する購買データと訪問場所データを組み合わせることで、どのようなファッションの好みを持つ人がどのような場所に出かけるのかという関係発見が期待できる。各データのパターンを抽出するだけでなく、各データのパターン間の関係も発見可能となる点に特徴がある。

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)

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
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
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
Masahiro KOHJIMA, Tatsushi MATSUBAYASHI, 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.6 pp.543-550 2016年6月
©電子情報通信学会 2016

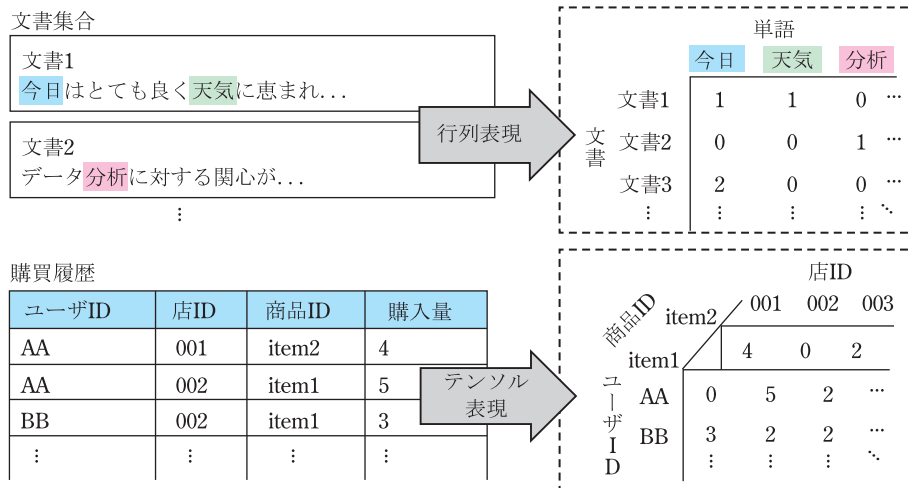


図1 解析対象データの例

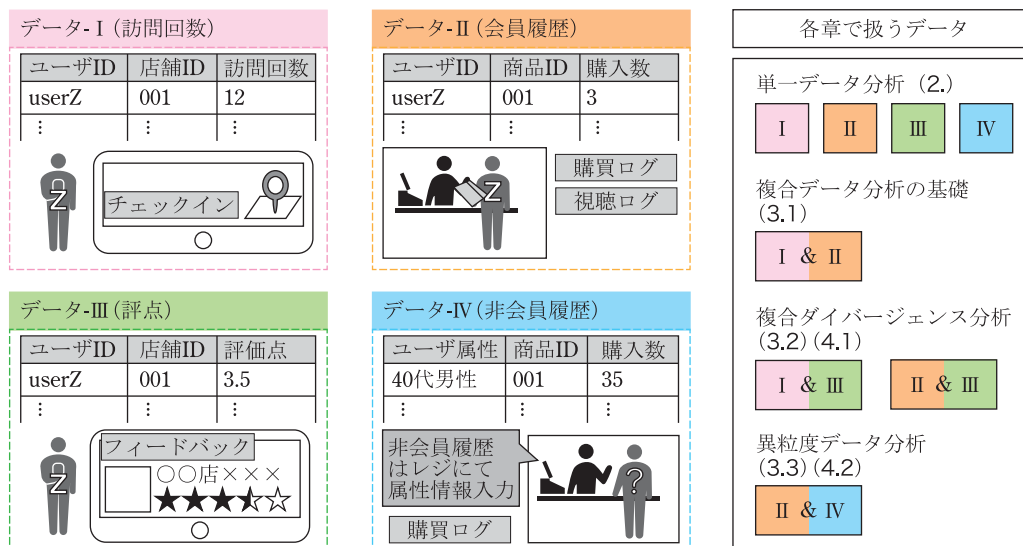


図2 データと対応する分析技術

連載第1回目の本稿は、NMFに基づく複合データ分析技術について紹介し、NTFについては第2回目の連載に譲る。本稿は次のような構成から成る。まず2.で、NMFを用いてある一つのデータを分析する単一データ分析について述べる。単一データ分析によって、図2に示すような訪問回数データ-I、購買データ-II、評点データ-IIIなどをそれぞれ単独で分析することが可能になる。3.では、データの組合せに応じて複数存在する複合データ分析手法について述べる。訪問・購買データ-I & IIを組み合わせて分析できる基礎的手法(3.1)、訪問・評点データの組-I & IIIなどを分析する混合ダイバージェンス分析手法(3.2)、会員購買・非会員購買データ-II & IVを分析する異粒度分析手法(3.3)

という複数の手法について幅広く紹介する。4.で複合データ分析手法の適用例を示す。

2. NMFによる単一データ分析

はじめにNMFの入出力を述べる。NMFの入力は要素が非負の行列 X である。以後行列 X は I 行 J 列の行列であるとし、その要素を x_{ij} と書く。また行列の全要素が0以上であることを $X \geq 0$ と書く。NMFの出力は、図3に示す X の因子分解の結果である I 行 R 列の因子行列 A と J 行 R 列の因子行列 B である。これらの行列の要素 a_{ir} , b_{jr} も非負であり、 $A, B \geq 0$ である。なお、 R は分解の因子数を表し、NMFの適用前に設定す

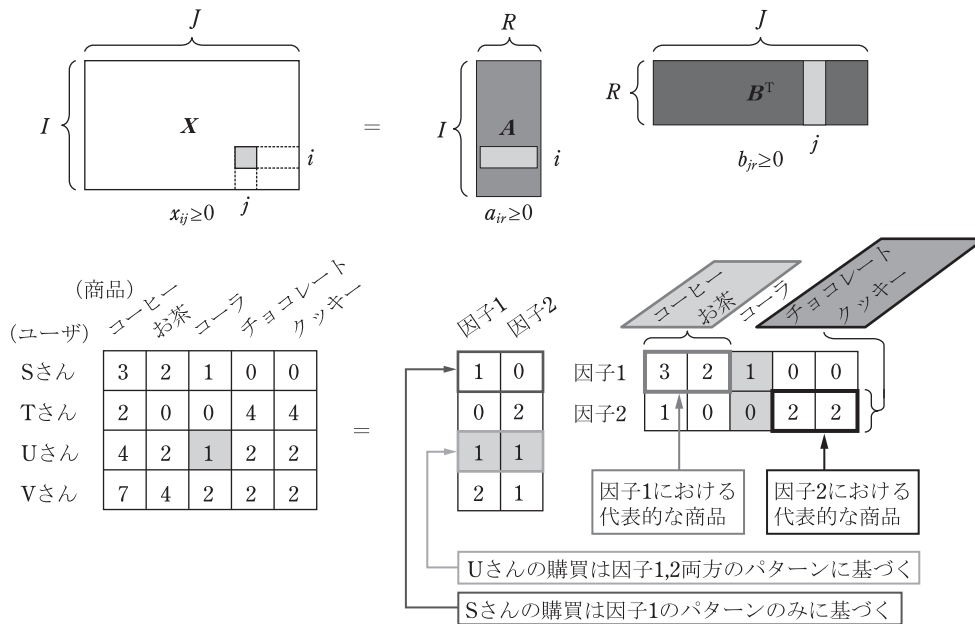


図3 NMFによる分解結果の例

る値である。

NMFの出力が抽出したパターンに対応する。図3下部の購買データでは、行列 B が各パターンの代表的な商品、行列 A が各ユーザの購買がどのパターンにどの程度基づくかを示す。行列 B の各因子で値の大きい代表商品は、因子1で「コーヒー」と「お茶」、因子2で「チョコレート」と「クッキー」である。よって「コーヒー好き」、「お菓子好き」の二つのパターンを抽出したと分かる。行列 A から、「Sさん」は因子1、「Uさん」は因子1, 2両方のパターンに基づく購買と分かる。

NMFは内部で最適化問題を解くことで因子行列を出力している。因子行列の積 AB^T を \hat{X} 、その要素を \hat{x}_{ij} と書く。図3の例では $X = \hat{X}$ が成立していたが、一般には等号を成立できない。そこでNMFでは、行列 X と \hat{X} の距離 $\mathcal{D}(X \parallel \hat{X})$ を定義し、この距離を最小化することで分解結果を得る。これは次の最適化問題として書き表せる。

$$\arg \min_{A, B} \mathcal{D}(X \parallel \hat{X}) \quad \text{s.t. } A, B \geq 0. \quad (1)$$

利用する行列間の距離には幾つかの候補が存在する。広く利用されるものには、要素間の距離 d に、ユークリッド距離の二乗を用いた d_{EU} や、一般化カルバックライブラ (KL) ダイバージェンスを利用した d_{KL} がある。

$$d_{\text{EU}}(x_{ij} \parallel \hat{x}_{ij}) = \frac{1}{2} (x_{ij} - \hat{x}_{ij})^2,$$

$$d_{\text{KL}}(x_{ij} \parallel \hat{x}_{ij}) = x_{ij} \log \frac{x_{ij}}{\hat{x}_{ij}} - x_{ij} + \hat{x}_{ij}.$$

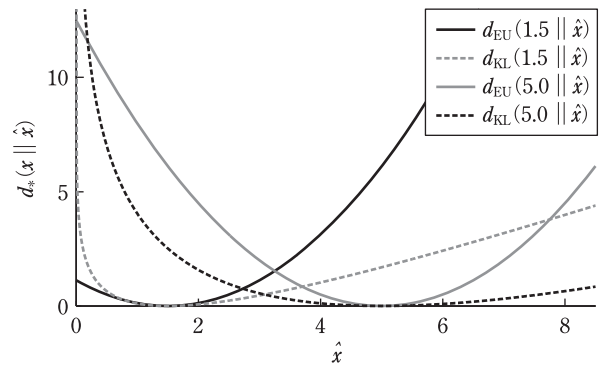


図4 ユークリッド距離とKLダイバージェンスの比較

これらの全要素に関する和をとることで行列間の距離を定義できる (記号*にはEUまたはKLが入る)。

$$\mathcal{D}_*(X \parallel \hat{X}) = \sum_{i,j=1}^{I,J} d_*(x_{ij} \parallel \hat{x}_{ij}). \quad (2)$$

距離による特性の違いを図4を用いて説明する。図4は $x=1.5, 5.0$ における各距離の値を示したものである。距離0となる点を中心に対称なユークリッド距離に対し、KLダイバージェンスは非対称で大きく見積もる ($\hat{x} \geq x$) ときの距離が、小さく見積もるときの距離より増加幅が小さくなる。また二つの距離を比較すると、 x の値が相対的に大きいとき ($x=5.0$) には横軸の広い範囲でKLダイバージェンスの方がユークリッド距離よりも値が小さく「鈍感」なものとなる。よって、行列要素が訪問回数などの大きい値を含む離散値となる場合、過剰なフィッティングを避けるためにKLダイバージェンス

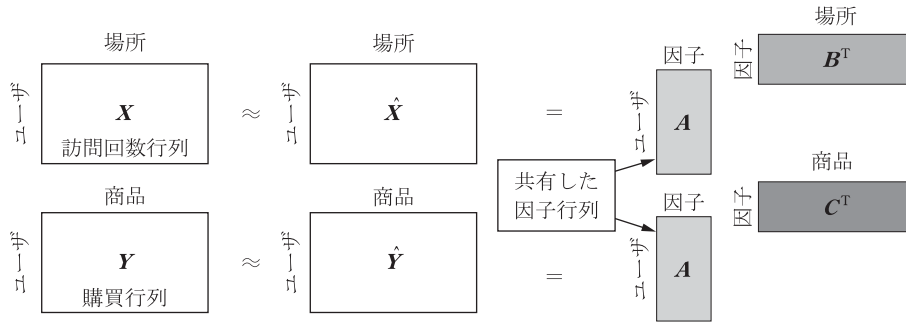


図5 因子行列を共有させた複数の行列の同時因子分解

を利用し、ある決まった範囲の値をとる映画の評点などの実数値となる場合、ユークリッド距離を用いる場合が多い。

式(1)を解くアルゴリズムは複数存在するが、ここでは実装上の簡易さから利用されることの多い乗法更新則に基づくアルゴリズムを紹介する。このアルゴリズムは、行列 \mathbf{A} , \mathbf{B} をランダムな非負値で初期化したのち、次の更新式に従い交互に \mathbf{A} , \mathbf{B} の更新を繰り返すことで分解結果を得るものである。

- ユークリッド距離を用いた場合：

$$a_{ir} \leftarrow a_{ir} \frac{\sum_{j=1}^J x_{ij} b_{jr}}{\sum_{j=1}^J \hat{x}_{ij} b_{jr}}, \quad b_{jr} \leftarrow b_{jr} \frac{\sum_{i=1}^I x_{ij} a_{ir}}{\sum_{i=1}^I \hat{x}_{ij} a_{ir}}.$$

- KL ダイバージェンスを用いた場合：

$$a_{ir} \leftarrow a_{ir} \frac{\sum_{j=1}^J \frac{x_{ij}}{\hat{x}_{ij}} b_{jr}}{\sum_{j=1}^J b_{jr}}, \quad b_{jr} \leftarrow b_{jr} \frac{\sum_{i=1}^I \frac{x_{ij}}{\hat{x}_{ij}} a_{ir}}{\sum_{i=1}^I a_{ir}}.$$

右辺に更新前の値を掛けていることが、この更新則が“乗法”更新則と呼ばれるゆえんである。このアルゴリズムは局所最適解への収束を保証する。全ての要素で等号 $x_{ij} = \hat{x}_{ij}$ が成り立つとき、右辺と左辺が等しくなり更新が停止することが容易に確かめられる。なお、 \mathbf{A} , \mathbf{B} の初期値を非負値に設定することで、更新式の右辺は常に非負の値となるため、収束先が非負値となることは保証される。なお更新則の導出は割愛したが、興味のある読者は文献(4)を参照してほしい。

3. 複合データ分析技術

3.1 複合データ分析の基礎

本章では、複合データ分析技術の紹介を行う。単一データ分析とは異なり、複合データ分析技術^{(5),(6)}では複数の行列を入力として利用する。本稿では入力行列の

数は二つであるとし、行列 \mathbf{X} 、行列 \mathbf{Y} をそれぞれ、行の指し示す事柄（ユーザなど）の等しい I 行 J 列、 I 行 K 列の行列であるとする。^(注1)これは例えば、図2中のユーザIDが二つのデータでひも付いている訪問データ-Iと会員購買データ-IIから、要素 x_{ij} がユーザ i の場所 j の訪問回数を表す行列 \mathbf{X} と、要素 y_{ik} がユーザ i の商品 k の購入数を表す行列 \mathbf{Y} を作成した場合に対応する。

複合データ分析の基礎となる手法は、この複数の入力行列を同時に因子分解する手法である(図5)。この手法の注目すべき点は、ユーザがどのパターンに基づくかを示す因子行列 \mathbf{A} を、二つの行列の因子分解に共通して用いる点にある。これにより、因子の等しいパターンの間に対応関係が導入される。例えば、図5に示す訪問・購買データへ適用することで、因子1のパターンに強く基づくユーザは、因子行列 \mathbf{B} の因子1の訪問パターンに基づいて、スポーツジムを代表的な場所として訪問し、因子行列 \mathbf{C} の因子1の購買パターンに基づいて、スポーツウェアを代表的な商品として購入する、といった結果が得られる。したがって、訪問パターンと購買パターンの間の関係発見が可能となる。

最適化問題としての定式化は非常に直感的である。非負値制約 $\mathbf{A}, \mathbf{B}, \mathbf{C} \geq 0$ に加えてユークリッド距離を用いる場合は式(3)、一般化KLダイバージェンスを用いる場合は式(4)で与えられる最適化問題を考える。

$$\arg \min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \{ \mathcal{D}_{\text{EU}}(\mathbf{X} \| \hat{\mathbf{X}}) + \mathcal{D}_{\text{EU}}(\mathbf{Y} \| \hat{\mathbf{Y}}) \}. \quad (3)$$

$$\arg \min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \{ \mathcal{D}_{\text{KL}}(\mathbf{X} \| \hat{\mathbf{X}}) + \mathcal{D}_{\text{KL}}(\mathbf{Y} \| \hat{\mathbf{Y}}) \}. \quad (4)$$

上記のように定式化された問題を解く乗法更新則は以下で与えられる。

- ユークリッド距離を用いた場合(式(3))：

(注1) 複数の行列、テンソルの因子分解手法には複数の名称が存在する。代表的なものとして、Collective Matrix Factorization⁽⁷⁾や Group Matrix Factorization⁽⁸⁾、Multiple Matrix/Tensor Factorization^{(5),(6)}、Coupled Tensor Factorization⁽⁹⁾が挙げられる。

$$a_{ir} \leftarrow a_{ir} \frac{\sum_{j=1}^J x_{ij} b_{jr} + \sum_{k=1}^K y_{ik} c_{kr}}{\sum_{j=1}^J \hat{x}_{ij} b_{jr} + \sum_{k=1}^K \hat{y}_{ik} c_{kr}},$$

$$b_{jr} \leftarrow b_{jr} \frac{\sum_{i=1}^I x_{ij} a_{ir}}{\sum_{i=1}^I \hat{x}_{ij} a_{ir}}, \quad c_{kr} \leftarrow c_{kr} \frac{\sum_{i=1}^I y_{ik} a_{ir}}{\sum_{i=1}^I \hat{y}_{ik} a_{ir}}.$$

$$a_{ir} \leftarrow a_{ir} \frac{-\beta_{ir} + \sqrt{\beta_{ir}^2 + 4(\sum_{j=1}^J \hat{x}_{ij} b_{jr}) \left(\sum_{k=1}^K \frac{y_{ik}}{\hat{y}_{ik}} c_{kr} \right)}}{2 \sum_{j=1}^J \hat{x}_{ij} b_{jr}}.$$

- KL ダイバージェンスを用いた場合 (式(4)):

$$a_{ir} \leftarrow a_{ir} \frac{\sum_{j=1}^J \frac{x_{ij}}{\hat{x}_{ij}} b_{jr} + \sum_{k=1}^K \frac{y_{ik}}{\hat{y}_{ik}} c_{kr}}{\sum_{j=1}^J b_{jr} + \sum_{k=1}^K c_{kr}},$$

$$b_{jr} \leftarrow b_{jr} \frac{\sum_{i=1}^I \frac{x_{ij}}{\hat{x}_{ij}} a_{ir}}{\sum_{i=1}^I a_{ir}}, \quad c_{kr} \leftarrow c_{kr} \frac{\sum_{i=1}^I \frac{y_{ik}}{\hat{y}_{ik}} a_{ir}}{\sum_{i=1}^I a_{ir}}.$$

いずれの距離を用いた場合も、因子行列 \mathbf{B} と \mathbf{C} の更新式は NMF の更新式と等しい。一方、共有された因子行列 \mathbf{A} の更新式には、行列 \mathbf{X}, \mathbf{Y} 両方を用いて計算される項が存在し、二つの行列へのフィッティングを考慮して値が推定されることが分かる。

3.2 混合ダイバージェンス分析

次に、行列ごとに異なる距離を用いた分析技術を紹介する⁽¹⁰⁾。この技術は、行列 \mathbf{X} の因子分解にはユークリッド距離、行列 \mathbf{Y} の因子分解に一般化 KL ダイバージェンスを用いて二つの行列を同時に因子分解する方法である。これは例えば図2における評点データ-Ⅲを表す行列 \mathbf{X} 、訪問回数データ-Ⅰを表す行列 \mathbf{Y} を分析する場合に現れる設定である。評点データと訪問回数データでは、行列要素のとり得る値の範囲が大きく異なるため、それぞれの行列で適切な距離を用いることが必要となる。図5と同じ因子分解を考えることで次のように定式化される。

$$\arg \min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \{ \mathcal{D}_{\text{EU}}(\mathbf{X} \| \hat{\mathbf{X}}) + \mathcal{D}_{\text{KL}}(\mathbf{Y} \| \hat{\mathbf{Y}}) \} \quad (5)$$

この場合の行列 \mathbf{A} の更新式は NMF と同様の導出により次のように求められる。

ただし、 $\beta_{ir} = \sum_{k=1}^K c_{kr} - \sum_{j=1}^J x_{ij} b_{jr}$ である。この更新式の形は導出の過程で二次方程式の解の公式を用いたことに由来している。なお、行列 \mathbf{B}, \mathbf{C} の更新式は前節におけるユークリッド距離、KL ダイバージェンスを用いた際の更新式と等しい。この技術によって、どの店舗を多く訪問するユーザがどの店舗を高く評価するかといった、店舗の訪問パターンと評価パターンの関係発見が期待できる。

3.3 異粒度データ分析

最後に二つの行列の行の対応関係の粒度が異なる場合の分析手法を紹介する⁽¹¹⁾。これは例えば図2で示した、ユーザ個人単位で収集される会員購買データ-Ⅱとユーザ属性単位で収集される非会員購買データ-Ⅳを分析する場合に現れる設定である。会員データ、非会員データをそれぞれ \mathbf{X}, \mathbf{Y} と書く。また、会員ユーザの属性情報が利用できるとし、それを行列 \mathbf{V} で表現する (ユーザ i が属性 k のとき $v_{ik}=1$, そうでなければ $v_{ik}=0$)。このような異粒度データの分析においては、行列 \mathbf{V} を因子行列間の制約条件に導入した図6に示す因子分解が有効であることが確認されている。最適化問題としての定式化と更新則は次のとおりである。

$$\arg \min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \{ \mathcal{D}_{\text{KL}}(\mathbf{X} \| \hat{\mathbf{X}}) + \mathcal{D}_{\text{KL}}(\mathbf{Y} \| \hat{\mathbf{Y}}) \} \quad \text{s.t. } \mathbf{V}^T \mathbf{A} = \mathbf{C}.$$

$$a_{ir} \leftarrow a_{ir} \frac{\left(\sum_{j=1}^J \frac{x_{ij}}{\hat{x}_{ij}} b_{jr} + \sum_{k=1}^K \sum_{j=1}^J v_{ik} \frac{y_{kj}}{\hat{y}_{kj}} b_{jr} \right)}{\sum_{j=1}^J b_{jr} + \sum_{k=1}^K \sum_{j=1}^J v_{ik} b_{jr}},$$

$$b_{jr} \leftarrow b_{jr} \frac{\left(\sum_{i=1}^I \frac{x_{ij}}{\hat{x}_{ij}} a_{ir} + \sum_{k=1}^K \frac{y_{kj}}{\hat{y}_{kj}} c_{kr} \right)}{\sum_{i=1}^I a_{ir} + \sum_{k=1}^K c_{kr}},$$

$$c_{kr} \leftarrow \sum_{i=1}^I v_{ik} a_{ir}.$$

この技術によって、会員ユーザと非会員ユーザの両方の購買を説明するパターンの抽出が可能となる。

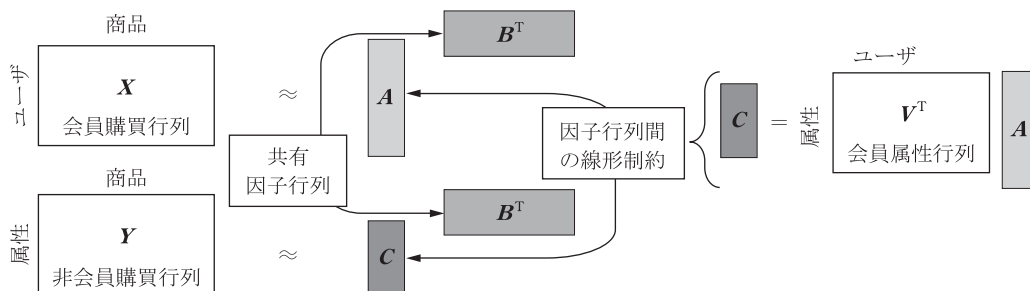
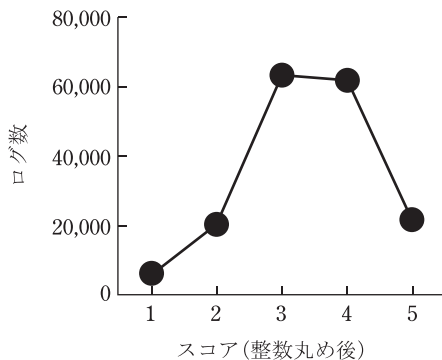


図6 異粒度データ分析⁽¹¹⁾における同時因子分解

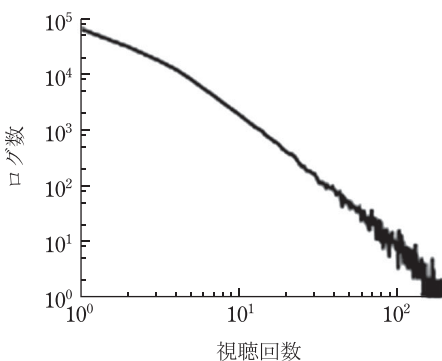
4. 適用例

4.1 映画カテゴリーの評価と視聴のパターン関係発見

はじめに混合ダイバージェンス分析 (3.2) の適用例として、MovieLens1M (ML) (注2) を利用した映画における評価と視聴に関するパターン間の関係発見の例を示す。ML はユーザの映画に対する 1~5 の値をとる評価スコアのデータである。このデータには映画のジャンル情報 (アクション, コメディ等計 18 種類) が含まれており、各映画は一つまたは複数のジャンルに属する。本実験では、ジャンルのペアを一つの解析単位とみなしてこれをカテゴリーと呼び、(一つのカテゴリーが、“アクションかつコメディ”, “子供向けかつアニメーションかつ冒険” のように複数のジャンルに対応する), 行列 X がユーザの各カテゴリーに対する平均スコア, 行列 Y がユーザの各カテゴリーの視聴回数を表現するよう作成した。なお、カテゴリーは該当カテゴリーに属する映画が五つ以上存在するものを利用し、作成した行列のサイズは $I=6,040, J=K=75$, 行列要素の値の分布は図7の



(a) 行列 X の要素の値の分布



(b) 行列 Y の要素の値の分布

図7 実験 (ML) における行列要素の値の分布

とおりである。行列 X の要素の値は連続値, 行列 Y の要素の値は離散かつ裾の長い分布であることから、今回は X の距離基準にユークリッド距離, Y には一般化 KL ダイバージェンスを利用する。

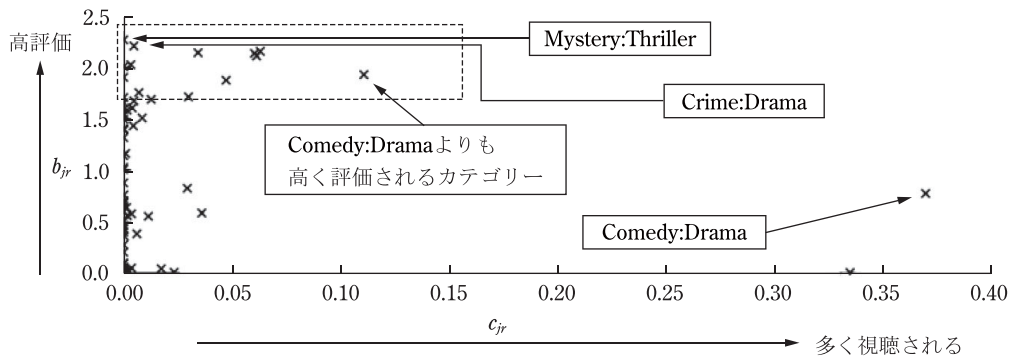
提案手法において因子数を $R=10$ と設定して得た因子のうち二つを図8に示す。散布図の縦軸はカテゴリー j の評点に関する因子 b_{jr} , 横軸は訪問に関する因子 c_{jr} である。図8 ($r=1$) に注目すると、Comedy: Drama が最も多く視聴されるカテゴリーだが、それよりも高い評価を得る Mystery: Thriller や Crime: Drama などが存在する。ここから、Comedy: Drama を視聴する習慣を持ちながらも、Mystery: Thriller の作品をより好むユーザが存在すると分かる。一方、図8 ($r=2$) に注目すると、Action: Thriller が散布図中で最も右上に位置している。また、Action: Adventure: Sci-Fi, Action: Crime: Drama も視聴回数が多くかつ高い評価を得ている。よって、熱狂的なアクションファンが存在すると分かる。このように混合ダイバージェンス分析によって、評価と視聴回数のパターンの関係が発見できることが分かる。

4.2 異粒度データからのパターン抽出

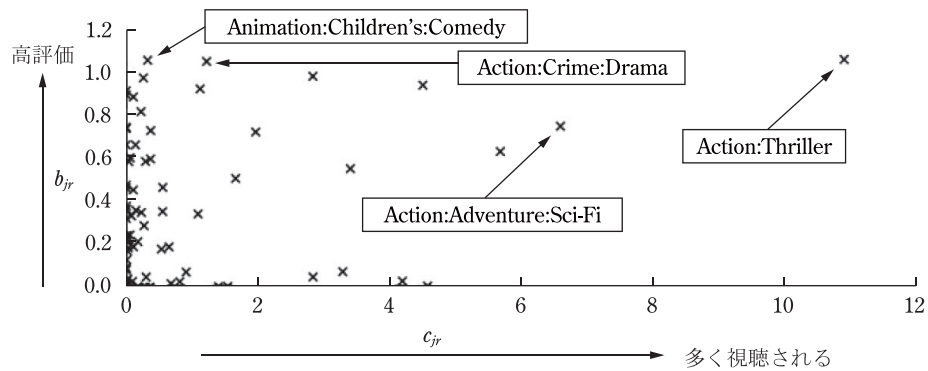
次に異粒度データ分析技術の基礎手法 (3.3) の適用例として、インテージ社の調査パネルデータ “SCI” を利用した、購買パターンの抽出を試みる。このうち1月についてはユーザ個人単位のデータは利用不可能であるという状況を想定し、2月の購買データからユーザごとの購買を表す行列 X , 1月の購買データからユーザ ID 情報を除去した、ユーザ属性ごとの購買を表す行列 Y を作成した。なお、両行列において要素の値は全て離散値であることから、今回は X, Y の両方の距離基準として一般化 KL ダイバージェンスを用いた異粒度データ分析手法を適用した。

図9に因子数 $R=10$ と設定して抽出した因子三つを示す。ヒストグラムの横軸が各因子 r における b_{jr}, c_{jr} の値の大きさ, 縦軸が値の大きい上位 10 件の商品, ユーザ属性を示す。図9 ($r=1$) では、“女性 35~49 代” や “女性 (>50)”, “主婦” が “ヨーグルト” や “牛乳” を多く購入することを示すパターンが抽出されている。これにより主婦層のユーザが日常的に乳製品を購入していることが分かる。また図9 ($r=2$) では、“男性 35~49 代”, “フルタイム社員” が “コーヒー” や “たばこ” などの嗜好品を多く購入することを示す「コーヒー好き」の購買パターンが抽出されている。これにより、例えば仕事の合間にコーヒーを購入する習慣を持つ、会社員などとして働くユーザが一定数存在すると推測できる。最後に図9 ($r=3$) では、“女性 35~49 代”, “女性 40 代” が “コーラ” や “スナック” などの菓子製品を購入することを示す「お菓子好き」の購買パターンが抽出されて

(注2) <http://www.grouplens.org/node/73>



(a) 基底 $r=1$ から作成した散布図



(b) 基底 $r=2$ から作成した散布図

図8 混合ダイバージェンス分析により得られた因子

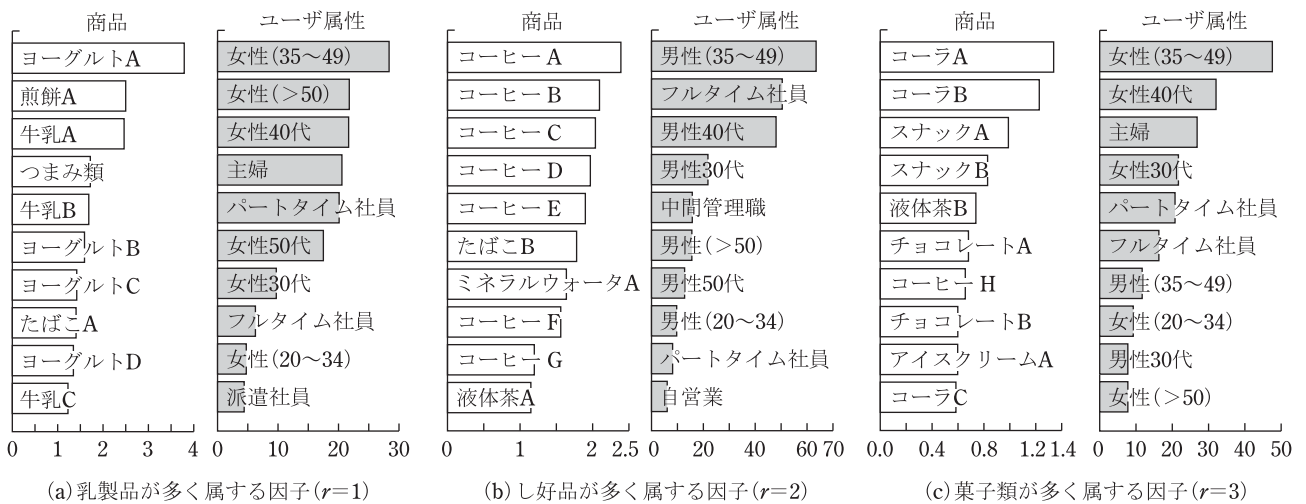


図9 異粒度データ分析により得られた因子

いる。女性に関する属性が多い点は因子($r=1$)のパターンと類似するが、相対的に“女性(>50)”の値が小さいことや“男性 35~49代”という属性が存在することも確認でき、「お菓子好き」パターンの購買が多い性別や年代を把握できる。このように異粒度データ分析

によって、ユーザ単位・属性単位それぞれのデータの分析、解釈が必要な従来のアプローチと異なり、粒度の違うデータのパターンを一度に抽出できることが分かる。

5. ま と め

本稿では、非負値行列分解に基づく複合データ分析技術について紹介した。昨今、複合データ分析技術は観光客分析などデータ分析の現場においても使われるようになってきている⁽¹²⁾。本稿を通して複合データ分析技術に関する興味、関心を持って頂ければ幸いである。NMFに関するより丁寧な説明が必要な読者には、解説記事⁽⁴⁾が助けになる。NMFのアルゴリズムの網羅的な把握には文献(13)を参照されたい。

文 献

- (1) D.D. Lee and H.S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788-791, 1999.
- (2) D.D. Lee and H.S. Seung, "Algorithms for non-negative matrix factorization," *Advances in neural information processing systems*, pp. 556-562, 2001.
- (3) A. Shashua and T. Hazan, "Non-negative tensor factorization with applications to statistics and computer vision," *Proceedings of the 22nd international conference on Machine learning*, pp. 792-799, 2005.
- (4) 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.
- (5) K. Takeuchi, K. Ishiguro, A. Kimura, and H. Sawada, "Non-negative multiple matrix factorization," *Proceedings of the 23rd international joint conference on Artificial Intelligence*, pp. 1713-1720, 2013.
- (6) K. Takeuchi, R. Tomioka, K. Ishiguro, A. Kimura, and H. Sawada, "Non-negative multiple tensor factorization," *IEEE 13th International Conference on Data Mining*, pp. 1199-1204, 2013.
- (7) A.P. Singh and G.J. Gordon, "Relational learning via collective matrix factorization," *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 650-658, 2008.
- (8) H. Lee and S. Choi, "Group nonnegative matrix factorization for EEG classification," *International Conference on Artificial Intelligence and Statistics*, pp. 320-327, 2009.
- (9) E. Acar, T.G. Kolda, and D.M. Dunlavy, "All-at-once optimization for coupled matrix and tensor factorizations," *KDD Workshop on Mining*

and Learning with Graphs, 2011.

- (10) 幸島匡宏, 松林達史, 澤田 宏, "訪問・評点データ統合解析のための非負値行列分解の検討," *信学技報, IBISML 2015-4*, pp. 21-26, June 2015.
- (11) M. Kohjima, T. Matsubayashi, and H. Sawada, "Probabilistic non-negative inconsistent-resolution matrices factorization," *Proceedings of the 24th ACM international conference on Information and Knowledge Management*, pp. 1855-1858, 2015.
- (12) 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.
- (13) A. Cichocki, R. Zdunek, A.H. Phan, and S. 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 1 日受付 平成 27 年 12 月 18 日最終受付)



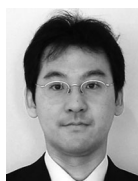
こうじま まさひろ
幸島 匡宏

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。



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。



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。