

非負制約下における複合行列分解と そのソーシャルメディア解析への応用

竹内 孝^{1,a)} 石黒 勝彦^{1,b)} 木村 昭悟^{1,c)} 澤田 宏^{2,d)}

受付日 2013年4月28日, 再受付日 2013年6月14日/2013年9月11日,
採録日 2013年10月8日

概要: 行列分解には、観測行列に含まれる零要素の割合が大きくなるにつれて低ランク近似の汎化性能が低下する問題がある。本稿では、この問題を解決するための統計的機械学習アプローチとして複合非負値行列因子分解 (Non-negative Multiple Matrix Factorization: NM2F) を提案する。NM2F は、観測行列と2つの補助行列の間に共通の潜在構造を仮定し、これらの行列を同時に分解する。本稿では、NM2F を非負値行列因子分解 (Non-negative Matrix Factorization: NMF) の一般化として定式化し、補助関数法により一般化 KL ダイバージェンスを用いた場合のパラメータ推定法を示す。さらに NM2F は、ブロック未定義領域あり NMF とポアソン分布を用いた確率的生成モデルと等価であることを示す。人工データと実データを用いた実験から、NM2F と既存手法の汎化性能を比較し、NM2F の定量的な優位性を示す。また、実データを用いた実験では、NM2F が複数の行列から多角的な基底を抽出する定性的な利点を示す。

キーワード: 機械学習, データマイニング, 行列分解

Multiple Matrix Factorization under the Non-negative Constraints and its Applications for Social Media Analysis

KOH TAKEUCHI^{1,a)} KATSUHIKO ISHIGURO^{1,b)} AKISATO KIMURA^{1,c)} HIROSHI SAWADA^{2,d)}

Received: April 28, 2013, Revised: June 14, 2013/September 11, 2013,
Accepted: October 8, 2013

Abstract: Analyzing highly sparse data often results in poor generalizing performances of matrix factorization. To compensate data sparseness, in this paper, we introduce a novel machine learning technique called Non-negative Multiple Matrix Factorization (NM2F). NM2F factorizes multiple matrices simultaneously under a non-negative constraint. We formulate NM2F as a generalization of Non-negative Matrix Factorization (NMF) with the generalized Kullback-Leibler divergence. We derive multiplicative update rules for parameter estimation. We evaluate NM2F and other existing techniques in both the quantitative and qualitative ways. NM2F shows better performance than other techniques on both synthetic and real-world data sets.

Keywords: machine learning, data mining, Matrix Factorization

1. はじめに

非負値行列因子分解 (Non-negative Matrix Factoriza-

¹ NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories, Kyoto 619-0237, Japan

² NTT サービスエボリューション研究所
NTT Service Evolution Laboratories, Yokosuka, Kanagawa 239-0847, Japan

a) takeuchi.koh@lab.ntt.co.jp

b) ishiguro.katsuhiko@lab.ntt.co.jp

c) akisato@ieee.org

d) sawada.hiroshi@lab.ntt.co.jp

tion: NMF) [21] は行列分解の1つである。その応用範囲は広く、音声響信号処理 [34]、画像処理 [19], [21]、テキストマイニング [38]、Brain Computer Interface [9] など多岐にわたる。NMF の非負制約は行列分解においてスパース制約として働くため、NMF による行列分解は、直感的に理解しやすい結果になると報告されている [12], [21]。また、非負制約に加えて他のスパース制約を課した NMF [7], [19] も提案されており、追加されたスパース制約によって、NMF は雑音の多いデータの分解においても良好な汎化性能を示している [1], [10], [18]。しかし、一般的に NMF を含む行

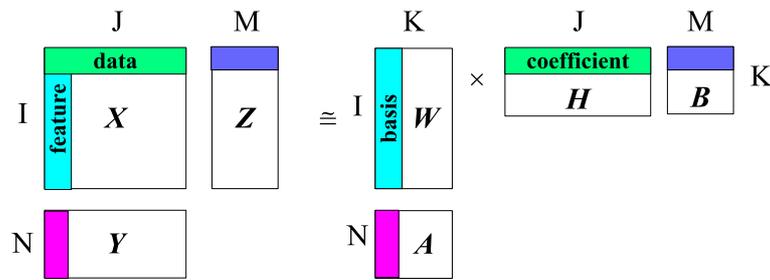


図 1 観測行列 X を基底行列 W と係数行列 H に分解する. その際, 補助行列 Y と Z を同時に分解する

Fig. 1 The target matrix X is factorized to the bases W and the coefficients H with the row wise auxiliary matrix Y and the column wise auxiliary matrix Z .

列分解は, 観測行列が高度にスパースな場合に汎化性能の低い解析結果を示す傾向がある [2].

近年, 人間の活動や環境の状態が, 多種類のサービスやセンサによって行動履歴データや環境センシングデータなどの大規模なデータの集合として観測されている. これらのデータの解析に行列分解を適用できれば, 複雑なデータの解釈が容易になると考えられる. しかし, 個々のデータは高次元からなるため観測行列が高度にスパースとなりやすく, 行列分解の適応には困難をともなう [20], [23]. この問題を解決する手段として, 複数のデータの同時解析があげられる. これは, たとえばデータの集合はある事象をテキスト, 音声, 映像, 位置情報などとして観測しているため, データ集合は観測した事象に起因した関係性を持つと仮定して行うデータ解析である.

このアイデアは, 協調フィルタリング分野のアイテム推薦問題において, 行列分解法 [20] のアドホックな拡張として用いられている. 実際, いくつかの研究 [26], [30], [36] では観測行列のスパース性を解決するために, 解析対象の観測行列に関係する補助的な観測行列を利用しアイテム推薦の性能を向上させている. しかし, 補助的なデータの利用に加えて非負制約を課した行列分解は, 著者らの知る限り提案されていない. これまで説明してきたように非負制約は直感的に理解しやすい分解結果を得るために重要である. また, 複合データ解析において非負制約がどのような効果を表すのか検証する必要がある.

本稿では, 観測データと補助データからなる複合行列の非負制約下での分解手法として, 複合非負値行列因子分解 (Non-negative Multiple Matrix Factorization: NM2F) を提案する. なお, 観測データと補助データから作る行列を観測行列と補助行列とし, 補助行列は観測行列と行あるいは列を共有するものとする (図 1). NM2F は複合行列に潜在構造の共有を仮定し, 同時に分解する. 同時分解によって, 高度にスパースな観測行列の汎化性能の向上が期待される. 協調フィルタリングの既存研究と異なる点としては, NM2F は定式化において非負制約を明示的に課している点がある. したがって, 従来研究と比較してより直感

的に理解しやすい基底行列と係数行列が得られるものと期待できる. 本稿では, NM2F を NMF の一般化として定式化し局所最適解を求めるためのパラメータ推定法を導出する. さらに, 本稿では, NM2F を確率モデルとして解釈とブロック状の未定義領域を持つ NMF としての解釈を行う. 人工データと実データを用いた実験によって, NM2F は高度にスパースな行列の分解において良好な性能を持つことを示す. また NM2F の分解結果を定性的に解釈し, NM2F の定性的に優れた性能を示す.

本稿の構成は以下のとおりである. 2 章では関連研究に対する本稿の新規性を述べる. 3 章では NM2F を提案する. 4 章では NM2F と他の既存手法の性能を人工データと実データを用いた実験を行い比較する. 5 章では本稿の結論を述べる.

2. 関連研究

非負値行列因子分解 (Non-negative Matrix Factorization: NMF) [21] は教師なしの行列分解で, 非負制約下で観測データを基底の重み付き線形和で低ランク近似し再構成誤差を最小化する. NMF は, 非負制約によってコスト関数が非凸となる問題を持つが, 一方で基底と重みが非負値のため解釈がしやすくさらに非負制約によってスパースな解が得られるという好ましい特徴を持つ. NMF を拡張し非負テンソルを解析する手法も提案されており, CP 分解を行う Non-negative Tensor Factorization (NTF) や, Tucker 分解を行う Non-negative Tucker Decomposition (NTD) が存在する [9]. NMF をノンパラメトリックベイズ法で拡張した手法も提案されている [16], [28]. これ以外にもデータの非負性に着目し直感的に解釈しやすいパターンを抽出するアルゴリズム [5], [13], [25] や, 同一モーダルの複数の非負データを互いに紐付けて分解する手法も提案されている [14], [15]. 本稿で提案する NM2F は, これらの非負制約に関する研究に類する.

NMF で利用される距離基準には, ユークリッド距離, 一般化 KL ダイバージェンス, 板倉斉藤距離などがある [9]. なお一般化 KL ダイバージェンスを用いた NMF

は、トピックモデル [3] の Probabilistic Latent Semantic Analysis (pLSA) [17] と等価である [11]. NMF のモデル推定法には、補助関数法 [9], [22] や、Alternating Least Squares (ALS) [8], Gibbs Sampling 法 [33] を用いられている. NMF のパラメータ推定は分散並列化によって高速化が可能であり [24], オンライン学習の枠組みで定式化された NMF も提案されている [6], [37]. 本稿では、NM2F を NMF の一般化として定式化しパラメータ推定方法を補助関数法によって導出する. したがって、上記の高速化手法を自然に適応することが可能であるため提案手法である NM2F はデータサイズに対しスケールする. 各パラメータの推定には一部の観測データのみを必要とするため、実用上は個々のデータを保持する分散環境でのパラメータ推定も可能となる.

行列分解の拡張は、アイテム推薦の分野で複数提案されている. これらの手法は主に確率的行列分解 (Probabilistic Matrix Factorization: PMF) を拡張している. PMF はアイテムとユーザの観測値をガウス分布による生成モデルで記述し、観測行列の最大事後確率 (MAP) 解を学習する. PMF の拡張法は様々で、PMF と LDA を融合させたモデル [36] や、PMF とソーシャルネットワーク上のリンク解析手法を融合させたモデル [26], [27], [29] が存在する. 近年では上記のモデルを融合させたモデル [30] も提案されている. このほかにもテキスト解析の研究に関連したモデル [35] が提案されている. いずれの場合も補助データを用いたことによる推薦性能の改善が示唆されている. PMF のほかには、nuclear norm を複数行列へ適応できるように拡張した手法 [4] などが提案されている. 本稿で提案する NM2F は、これらの研究と深く関係している. しかし、非負制約条件を課する点で異なっている. また文献 [4] では、再構成誤差にユークリッド距離のみを仮定している点にも相違がある.

3. 複合非負値行列因子分解

3.1 問題の定式化

本稿では、1つの観測データと2つの補助データを扱う. 主観測データは I 次元からなり、データ点数を J とする. j 点目のデータにおける次元 i の観測値を $x_{i,j}$ とおき、 j 点目のデータの観測ベクトルを $\mathbf{x}_j = (x_{1,j}, \dots, x_{I,j})^T \in \mathbb{R}_+^I$ とする*1. 観測データは \mathbf{x}_j を列ベクトルに持つ観測行列 \mathbf{X} として、

$$\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_J) \in \mathbb{R}_+^{I \times J}, \quad (1)$$

と定める.

1つ目の補助データは、 \mathbf{X} の j 番目のデータに対応して観測される. 補助データの次元を N とする. j 点目の補

助データにおける次元 n の観測値を $y_{n,j}$ とおき、 j 点目の補助データの観測ベクトルを $\mathbf{y}_j = (y_{1,j}, \dots, y_{N,j})^T \in \mathbb{R}_+^N$ とする. 補助データは \mathbf{y}_j を列ベクトルに持つ補助行列 \mathbf{Y} として、

$$\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_J) \in \mathbb{R}_+^{N \times J}, \quad (2)$$

と定める.

2つ目の補助データは、 \mathbf{X} の i 次元に対応して観測される. 補助データの点数を M とする. m 点目の補助データにおける次元 i の観測値を $z_{i,m}$ とし、 m 点目の補助データの観測ベクトルを $\mathbf{z}_m = (z_{1,m}, \dots, z_{I,m})^T \in \mathbb{R}_+^I$ とする. 補助データは \mathbf{z}_m を列ベクトルに持つ補助行列 \mathbf{Z} として、

$$\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_M) \in \mathbb{R}_+^{I \times M}, \quad (3)$$

と定める. 行列 \mathbf{X} , \mathbf{Y} , \mathbf{Z} の関係を図 1 の左側に示す.

本稿の目的は、観測行列 \mathbf{X} からの基底行列 \mathbf{W} と係数行列 \mathbf{H} の推定である. 提案手法は補助行列 \mathbf{Y} と \mathbf{Z} を利用するため、既存手法と比較して汎化性能の向上が期待できる. さらに補助行列からも基底行列と係数行列を抽出するため、補助行列に潜在するパターンも同時に抽出できる利点を持つと期待できる.

観測行列と補助行列の具体例を説明する. 4章では、Last.fm という音楽視聴サービスのデータセットを利用する. このデータセットでは、観測行列 \mathbf{X} はあるユーザがアーティストの音楽を視聴した回数、補助行列 \mathbf{Y} はあるタグがアーティストに付与された回数、補助行列 \mathbf{Z} はあるユーザと別のユーザの友人関係情報である.

3.2 提案手法

NM2F は、分解後の行列がすべて非負となる制約の下で観測行列 \mathbf{X} と補助行列 \mathbf{Y} , \mathbf{Z} を分解する. 分解から得られる基底行列と係数行列は、非負制約によってスパースになり、さらに直感的に理解しやすいものとなると期待される. 我々のアイデアは、観測行列 \mathbf{X} が高度にスパースで行列分解が困難な際に補助行列を用いることで、利用できる情報量を増やすことである. NM2F は観測行列を分解するために2つの補助行列を利用することで、より性能の高い基底行列と係数行列の推定が可能になる.

観測行列は K 個の基底を持つと仮定する. 観測行列 \mathbf{X} の k 番目の基底ベクトルを $\mathbf{w}_k = (w_{1,k}, \dots, w_{I,k})^T \in \mathbb{R}_+^I$ とし、基底行列を $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_K) \in \mathbb{R}_+^{I \times K}$ とおく. 行方向の補助行列の k 番目の基底ベクトルを $\mathbf{a}_k = (a_{1,k}, \dots, a_{N,k})^T \in \mathbb{R}_+^N$ とし、補助行列 \mathbf{Y} の基底行列を $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_K) \in \mathbb{R}_+^{N \times K}$ とおく.

次に観測行列 \mathbf{X} の j 点目のデータに対応する係数ベクトルを $\mathbf{h}_j = (h_{1,j}, \dots, h_{K,j})^T \in \mathbb{R}_+^K$ とし、係数行列を $\mathbf{H} = (\mathbf{h}_1, \dots, \mathbf{h}_J) \in \mathbb{R}_+^{K \times J}$ と定める. 列方向の補助行列を \mathbf{Z} の m 点目のデータに対応する係数ベク

*1 本稿では、ベクトルを小文字の太字、行列を大文字の太字、転置を T で表す.

トル $\mathbf{b}_m = (b_{1,m}, \dots, h_{K,m})^T \in \mathbb{R}_+^K$ とし、係数行列を $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_M) \in \mathbb{R}_+^{K \times M}$ と定める。このとき、観測値 $x_{i,j}, y_{n,j}, z_{i,m}$ の再構成値 $\hat{x}_{i,j}, \hat{y}_{n,j}, \hat{z}_{i,m}$ は、基底と係数の重み付き線形和として、

$$\begin{aligned} \hat{x}_{i,j} &= \sum_{k=1}^K w_{i,k} h_{k,j}, \hat{y}_{n,j} = \sum_{k=1}^K a_{n,k} h_{k,j}, \\ \hat{z}_{i,m} &= \sum_{k=1}^K w_{i,k} b_{k,m}, \end{aligned} \quad (4)$$

と定める。

NM2F は以下の仮定をおいている。まず行方向の補助行列 \mathbf{Y} は観測行列 \mathbf{X} と基底行列 \mathbf{W} を共有すると仮定する。次に列方向の補助行列 \mathbf{Z} は観測行列 \mathbf{X} と係数行列 \mathbf{H} を共有すると仮定する。ここで $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ の再構成値を $\hat{\mathbf{X}} \in \mathbb{R}_+^{I \times J}, \hat{\mathbf{Y}} \in \mathbb{R}_+^{N \times J}, \hat{\mathbf{Z}} \in \mathbb{R}_+^{I \times M}$ とする。次に、 h_j, \mathbf{b}_m を \mathbb{R}_+^K 上の元とし、線形写像 $W: \mathbb{R}_+^K \xrightarrow{\mathbf{W}} \mathbb{R}_+^I, A: \mathbb{R}_+^K \xrightarrow{\mathbf{A}} \mathbb{R}_+^N$ を定める。すると、 $\hat{\mathbf{X}}, \hat{\mathbf{Y}}$ は共通の係数行列 \mathbf{H} を持ち、異なる線形写像 \mathbf{W}, \mathbf{A} によって得られる。また、 $\hat{\mathbf{X}}, \hat{\mathbf{Z}}$ は共通の線形写像 \mathbf{W} と、異なる係数行列 \mathbf{H}, \mathbf{B} によって得られると仮定をおいたことになる。ただし補助行列 \mathbf{Y} と \mathbf{Z} は観測行列と異なるスケールを持つとし、スケール変数 $\alpha \geq 0, \beta \geq 0$ によってスケールの差異をモデル化する。

NM2F は、行列 $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ を行列 $\mathbf{W}, \mathbf{A}, \mathbf{H}, \mathbf{B}$ で再構成した際の再構成誤差 $\mathcal{D}(\mathbf{X}, \mathbf{Y}, \mathbf{Z} | \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}; \alpha, \beta)$ を最小化する。

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}} \mathcal{D}(\mathbf{X}, \mathbf{Y}, \mathbf{Z} | \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}; \alpha, \beta) \\ \text{s. t. } \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B} \geq 0. \end{aligned} \quad (5)$$

このとき、 α, β は $\alpha \geq 0, \beta \geq 0$ を満たすスケール変数であり、再構成誤差 \mathcal{D} は、行列の要素 $x_{i,j}$ とその再構成値 $\hat{x}_{i,j}$ との間の誤差 $d(x_{i,j} | \hat{x}_{i,j})$ を用いて、

$$\begin{aligned} \mathcal{D}(\mathbf{X}, \mathbf{Y}, \mathbf{Z} | \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}; \alpha, \beta) &= \mathcal{D}(\mathbf{X} | \mathbf{W}, \mathbf{H}) + \alpha \mathcal{D}(\mathbf{Y} | \mathbf{A}, \mathbf{H}) + \beta \mathcal{D}(\mathbf{Z} | \mathbf{W}, \mathbf{B}) \\ &= \mathcal{D}(\mathbf{X} | \hat{\mathbf{X}}) + \alpha \mathcal{D}(\mathbf{Y} | \hat{\mathbf{Y}}) + \beta \mathcal{D}(\mathbf{Z} | \hat{\mathbf{Z}}) \\ &= \sum_{i=1}^I \sum_{j=1}^J d(x_{i,j} | \hat{x}_{i,j}) + \alpha \sum_{n=1}^N \sum_{j=1}^J d(y_{n,j} | \hat{y}_{n,j}) \\ &\quad + \beta \sum_{i=1}^I \sum_{m=1}^M d(z_{i,m} | \hat{z}_{i,m}), \end{aligned} \quad (6)$$

とする。

本稿では再構成誤差 \mathcal{D} に一般化 KL ダイバージェンス (generalized Kullback-Leibler divergence: gKL) を用いる。その際、行列の要素 $x_{i,j}$ とその再構成値 $\hat{x}_{i,j}$ の間の誤差を、

$$d_{gKL}(x_{i,j} | \hat{x}_{i,j}) = x_{i,j} \log \frac{x_{i,j}}{\hat{x}_{i,j}} - x_{i,j} + \hat{x}_{i,j}, \quad (7)$$

とすると、NM2F の一般化 KL ダイバージェンスは、

$$\begin{aligned} \mathcal{D}(\mathbf{X}, \mathbf{Y}, \mathbf{Z} | \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}; \alpha, \beta) &= \sum_{i=1}^I \sum_{j=1}^J d_{gKL}(x_{i,j} | \hat{x}_{i,j}) + \alpha \sum_{n=1}^N \sum_{j=1}^J d_{gKL}(y_{n,j} | \hat{y}_{n,j}) \\ &\quad + \beta \sum_{i=1}^I \sum_{m=1}^M d_{gKL}(z_{i,m} | \hat{z}_{i,m}). \end{aligned} \quad (8)$$

となる。

一般化 KL ダイバージェンスは NMF に広く用いられる再構成誤差である。NM2F は一般化 KL ダイバージェンスに限らず、ユークリッド距離や板倉齋藤距離を用いた定式化も可能である。なお、一般化 KL ダイバージェンスはある変数 $\lambda \geq 0$ を定めると、

$$\lambda d_{gKL}(x_{i,j} | \hat{x}_{i,j}) = d_{gKL}(\lambda x_{i,j} | \lambda \hat{x}_{i,j}), \quad (9)$$

が成り立つ。これは式 (18) で用いる。

3.3 NMF との比較

NMF は、観測行列 \mathbf{X} を基底行列 \mathbf{W} と係数行列 \mathbf{H} に分解し、再構成した際の再構成誤差 $\mathcal{D}(\mathbf{X} | \mathbf{W}, \mathbf{H})$ を最小化する。

$$\min_{\mathbf{W}, \mathbf{H}} \mathcal{D}(\mathbf{X} | \mathbf{W}, \mathbf{H}) \text{ s. t. } \mathbf{W}, \mathbf{H} \geq 0. \quad (10)$$

スケール変数 $\alpha = \beta = 0$ とするとき、NM2F の再構成誤差は、

$$\mathcal{D}(\mathbf{X}, \mathbf{Y}, \mathbf{Z} | \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}; \alpha, \beta) = \mathcal{D}(\mathbf{X} | \mathbf{W}, \mathbf{H}), \quad (11)$$

となる。したがって NM2F は NMF の一般化となっている。

3.4 モデル推定

次の乗法的更新則によって、再構成誤差 \mathcal{D} を最小化する最適な $\mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}$ を求められる。導出方法の詳細は、付録に示す。

$$\begin{aligned} w_{i,k}^{\text{new}} &= w_{i,k} \frac{\left(\sum_{j \in \mathcal{J}_i} \frac{x_{i,j}}{\hat{x}_{i,j}} h_{k,j} + \beta \sum_{m \in \mathcal{M}_i} \frac{z_{i,m}}{\hat{z}_{i,m}} b_{k,m} \right)}{\sum_j h_{k,j} + \beta \sum_m b_{k,m}}, \\ h_{k,j}^{\text{new}} &= h_{k,j} \frac{\left(\sum_{i \in \mathcal{I}_j} \frac{x_{i,j}}{\hat{x}_{i,j}} w_{k,j} + \alpha \sum_{n \in \mathcal{N}_j} \frac{y_{n,j}}{\hat{y}_{n,j}} a_{n,k} \right)}{\sum_i w_{k,j} + \alpha \sum_n a_{n,k}}, \\ a_{n,k}^{\text{new}} &= a_{n,k} \frac{\sum_{j \in \mathcal{J}_n} \frac{y_{n,j}}{\hat{y}_{n,j}} h_{k,j}}{\sum_j h_{k,j}}, \\ b_{k,m}^{\text{new}} &= b_{k,m} \frac{\sum_{i \in \mathcal{I}_m} \frac{z_{i,m}}{\hat{z}_{i,m}} w_{i,k}}{\sum_i w_{i,k}}. \end{aligned} \quad (12)$$

乗法的更新則の計算を効率的に行うために X の第 j 列ベクトルにおける非零要素の集合を \mathcal{I}_j とし、第 j 列ベクトルの和を $\sum_{i \in \mathcal{I}_j}$ と定める。同様に、 X の第 i 行ベクトル、 Y の第 j 列ベクトル、 Z の第 i 列ベクトルの非零要素の集合を \mathcal{J}_i , \mathcal{M}_i , \mathcal{N}_j とした。乗法的更新則においても、 $\alpha = 0$, $\beta = 0$ のとき、NM2F の更新則は NMF の更新則と一致することから、NM2F は NMF を一般化したモデルといえる。スケーリング変数 α , β は交差検定法によって決定する。NM2F の擬似コードを Algorithm1 に示す。

Algorithm 1 NM2F の擬似コード

Require: $X, Y, Z, K, \alpha, \beta, Iteration$
 一様分布からのサンプリングにより初期化: W, H, A, B
for $l = 1$ **to** $Iteration$ **do**
 for $i = 1$ **to** I **do**
 for $k = 1$ **to** K **do**
 $w_{i,k} \leftarrow w_{i,k} \frac{\left(\sum_{j \in \mathcal{I}_j} \frac{x_{i,j}}{\bar{x}_{i,j}} h_{k,j} + \beta \sum_{m \in \mathcal{M}_i} \frac{z_{k,m}}{\bar{z}_{k,m}} b_{i,m} \right)}{\sum_j h_{k,j} + \beta \sum_m b_{k,m}}$
 end for
 end for
 for $k = 1$ **to** K **do**
 for $j = 1$ **to** J **do**
 $h_{k,j} \leftarrow h_{k,j} \frac{\left(\sum_{i \in \mathcal{I}_j} \frac{x_{i,j}}{\bar{x}_{i,j}} w_{k,j} + \alpha \sum_{n \in \mathcal{N}_j} \frac{y_{n,j}}{\bar{y}_{n,j}} a_{n,k} \right)}{\sum_i w_{k,j} + \alpha \sum_n a_{n,k}}$
 end for
 end for
 for $n = 1$ **to** N **do**
 for $k = 1$ **to** K **do**
 $a_{n,k} \leftarrow a_{n,k} \frac{\sum_{j \in \mathcal{N}_j} \frac{y_{n,j}}{\bar{y}_{n,j}} h_{k,j}}{\sum_j h_{k,j}}$
 end for
 end for
 for $k = 1$ **to** K **do**
 for $m = 1$ **to** M **do**
 $b_{k,m} \leftarrow b_{k,m} \frac{\sum_{i \in \mathcal{M}_i} \frac{z_{i,m}}{\bar{z}_{i,m}} w_{i,k}}{\sum_i w_{i,k}}$
 end for
 end for
end for
return W, H, A, B

3.5 ブロック未定義領域 NMF としての解釈

NM2F は未定義領域のある NMF として解釈できる。まず、NMF によって観測行列 $T \in \mathbb{R}_+^{(I+N) \times (J+M)}$ を基底行列 $U \in \mathbb{R}_+^{(I+N) \times K}$ と係数行列 $V \in \mathbb{R}_+^{K \times (J+M)}$ へ分解する問題を考える。ただし T には未定義 (don't care) 領域が存在すると仮定する。未定義領域を示す変数として $\omega_{i,j}$ を、

$$\omega_{i,j} = \begin{cases} 0 & t_{i,j} \text{ は未定義領域,} \\ 1 & \text{otherwise} \end{cases} \quad (13)$$

と定め、 $\omega_{i,j}$ を要素に持つ行列を $\Omega \in \mathbb{R}^{(I+N) \times (J+M)}$ とする。ここでは特に、未定義領域が行列 T の右下部に集まる場合、

$$\omega_{i,j} = \begin{cases} 0 & i > I \text{ かつ } j > J \text{ なら未定義領域,} \\ 1 & \text{otherwise} \end{cases} \quad (14)$$

を考える。未定義領域あり NMF の再構成誤差 \hat{D} を、

$$\hat{D}(T, \Omega | U, V) = \sum_{i=1}^{I+N} \sum_{j=1}^{J+M} \omega_{i,j} d(t_{i,j} | \hat{t}_{i,j}), \quad (15)$$

とおく。このとき \hat{D} は、

$$\begin{aligned} \hat{D}(T, \Omega | T, V) &= \sum_{i=I}^I \sum_{j=J}^J 1 \times d(t_{i,j} | \hat{t}_{i,j}) + \sum_{i=I+1}^{I+N} \sum_{j=1}^J 1 \times d(t_{i,j} | \hat{t}_{i,j}) \\ &+ \sum_{i=1}^I \sum_{j=J+1}^{J+M} 1 \times d(t_{i,j} | \hat{t}_{i,j}) + \sum_{i=I+1}^{I+N} \sum_{j=J+1}^{J+M} 0 \times d(t_{i,j} | \hat{t}_{i,j}), \end{aligned} \quad (16)$$

と書き下せる。ここで T, U, V を、

$$T = \begin{pmatrix} X \hat{\beta} Z \\ \hat{\alpha} Y O \end{pmatrix}, U = \begin{pmatrix} W \\ \hat{\alpha} A \end{pmatrix}, V = (H \hat{\beta} B), \quad (17)$$

とおく。なお $O \in \mathbb{R}_+^{N \times M}$ は任意の行列、また $\hat{\alpha} = \frac{\alpha}{INK}$, $\hat{\beta} = \frac{\beta}{IMK}$ とした。すると \hat{D} は、

$$\begin{aligned} \hat{D}(T, \Omega | W, H) &= \sum_{i=I}^I \sum_{j=1}^J d(x_{i,j} | \hat{x}_{i,j}) + \alpha \sum_{n=1}^N \sum_{j=1}^J d(y_{n,j} | \hat{y}_{n,j}) \\ &+ \beta \sum_{i=1}^I \sum_{m=1}^M d(z_{i,m} | \hat{z}_{i,m}) \\ &= \mathcal{D}(X, Y, Z | W, H, A, B; \alpha, \beta), \end{aligned} \quad (18)$$

となり、 $\mathcal{D}(X, Y, Z | W, H, A, B; \alpha, \beta)$ と一致する。このとき式 (9) を用いてスケーリング変数をくり出ししている。未定義領域を欠損値とみなした場合は、欠損値あり NMF [31] を複数観測行列を分解できるように拡張するために、欠損値をブロック状に配置したものと解釈できる。

3.6 確率的生成モデルとしての解釈

NM2F は確率的生成モデルとしての解釈も可能である。一般化 KL ダイバージェンスを利用した NM2F は、NMF と同様に、ポアソン分布を用いた次の確率的生成モデルとして書き下せる。観測値 η が平均値 μ のポアソン分布、

$$p(\eta; \mu) = \frac{\mu^\eta}{\eta!} \exp(-\mu), \quad (19)$$

に従うとする。ここで上記のポアソン分布と、観測値が η で平均値が η のポアソン分布の対数尤度比を定めると、対数尤度比は負の一般化 KL ダイバージェンスにおける要素の誤差と一致する。

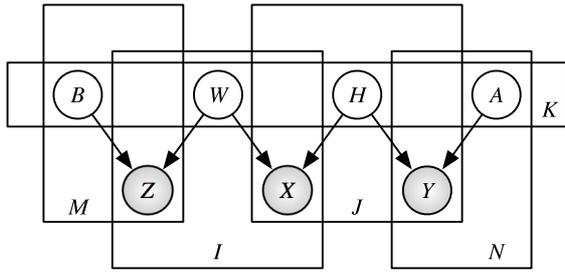


図 2 NM2F のグラフィカルモデル

Fig. 2 Graphical model of Non-negative Multiple Matrix Factorization. \mathbf{X} and \mathbf{Z} are generated from the same bases \mathbf{W} but different coefficients \mathbf{H} and \mathbf{B} . \mathbf{Y} is generated from bases \mathbf{A} and coefficients \mathbf{H} shared by \mathbf{X} .

$$\begin{aligned} \ln \frac{p(\eta; \mu)}{p(\eta; \eta)} &= \eta \ln \frac{\mu}{\eta} + \eta - \mu \\ &= -d_{gKL}(\eta | \mu). \end{aligned} \quad (20)$$

$p(\eta; \eta)$ は定数と見なせるため、対数尤度と対数尤度比は定数の差の範囲で近似的に一致する。

$$\begin{aligned} \ln \frac{p(\eta; \mu)}{p(\eta; \eta)} &= \ln p(\eta; \mu) - \ln p(\eta; \eta) \\ &= \ln p(\eta; \mu) + \text{const}, \end{aligned} \quad (21)$$

以上より、ポアソン分布の対数尤度は負の一般化 KL ダイバージェンスの要素の誤差と近似的に一致することが分かる。

NM2F の確率的生成モデルの対数尤度は、

$$\begin{aligned} &\ln p(\mathbf{X}, \hat{\alpha} \mathbf{Y}, \hat{\beta} \mathbf{Z} | \mathbf{W}, \mathbf{H}, \hat{\alpha} \mathbf{A}, \hat{\beta} \mathbf{B}) \\ &= \ln p(\mathbf{X}, | \mathbf{W}, \mathbf{H}) + \ln p(\hat{\alpha} \mathbf{Y} | \mathbf{H}, \hat{\alpha} \mathbf{A}) \\ &+ \ln p(\hat{\beta} \mathbf{Z} | \mathbf{W}, \hat{\beta} \mathbf{B}) \\ &= \sum_{i=1}^I \sum_{j=1}^J \ln p(x_{i,j} | \hat{x}_{i,j}) + \alpha \sum_{n=1}^N \sum_{j=1}^J \ln p(y_{n,j} | \hat{y}_{n,j}) \\ &+ \beta \sum_{i=1}^I \sum_{m=1}^M \ln p(z_{i,m} | \hat{z}_{i,m}) \\ &\simeq -\mathcal{D}(\mathbf{X}, \mathbf{Y}, \mathbf{Z} | \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}; \alpha, \beta), \end{aligned} \quad (22)$$

となる。したがって、式 (7) の最小化は、上記の確率的生成モデルの対数尤度最大化と近似的に等価である。

NM2F のグラフィカルモデルを図 2 に示す。グラフィカルモデルから、 \mathbf{X} , \mathbf{Y} , \mathbf{Z} が行列 \mathbf{W} , \mathbf{H} を共有する構造が確認できる。

NM2F を確率的生成モデルとして解釈することで、NM2F のモデルに従う人工データを確率的に生成 (サンプリング) 可能となる。具体的なデータ生成の過程を Algorithm2 に示す。

4. 実験

4.1 評価指標

モデルの性能を定量的に評価する指標として、データ

Algorithm 2 NM2F のデータ生成過程

Require: I, J, M, N, K

一様分布からのサンプリングにより初期化: $\mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}$

for $i = 1$ to I do

for $j = 1$ to J do

$x_{i,j} \sim \text{Poisson}(x_{i,j} | \sum_{k=1}^K w_{i,k} h_{k,j})$

end for

end for

for $n = 1$ to N do

for $j = 1$ to J do

$y_{n,j} \sim \text{Poisson}(y_{n,j} | \sum_{k=1}^K a_{n,k} h_{k,j})$

end for

end for

for $i = 1$ to I do

for $m = 1$ to M do

$z_{i,m} \sim \text{Poisson}(z_{i,m} | \sum_{k=1}^K w_{i,k} b_{k,m})$

end for

end for

return $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$

セットに対するモデルの平均対数尤度を用いる。観測データ点数 N の異なるデータセットを用いた実験間でのモデルの性能を比較するために、観測データ点ごとの対数尤度である平均対数尤度を採用した。実験では、与えられたデータセットをトレーニングセットとテストセットに分割した後、トレーニングセットに対する対数尤度を最大化するようモデルのパラメータの学習を行う。モデルの汎化性能を比較するために、テストセットに対する平均対数尤度を用いる。テストセットに対する平均対数尤度が高いモデルほど、データの潜在的な構造をとらえた良いモデルといえる。平均対数尤度 f を次のように定める。

$$f(\mathbf{X} | \theta) = \frac{1}{M} \sum_{m=1}^M \log p(x_m | \theta), \quad (23)$$

なお、テストセットのデータ点数を M とし、 θ はモデルの推定したパラメータである。具体的には、 θ は $\theta = \{\mathbf{W}, \mathbf{H}\}$ とする。実験では、観測行列 \mathbf{X} の非零観測要素を 5 つのセットに分割し、5 交差検定法によってテストセットの平均対数尤度の平均値と標準偏差を求める。

4.2 人工データ実験

人工データ実験では NMF, VBNMF (Variational Bayesian Non-negative Matrix Factorization) [7], PMF [32], そして NM2F の性能を比較する。従来手法は \mathbf{X} を分解対象とし、NM2F は \mathbf{X} , \mathbf{Y} , \mathbf{Z} を分解対象とする。従来手法と提案手法のデータセットに対する汎化性能を比較するために、 \mathbf{X} のテストセットに対する平均対数尤度を用いる。NM2F は補助行列 \mathbf{Y} , \mathbf{Z} を同時に分解するため、汎化性能の向上が期待される。

また、人工データは図 2 の確率的生成モデルによって生成する。このとき、 \mathbf{X} , \mathbf{Y} , \mathbf{A} , \mathbf{B} の次元は、 $I = J = N = M = 100$ とする。 \mathbf{W} , \mathbf{H} , \mathbf{A} , \mathbf{B} の基底数は、 $K = 10$ と

表 1 人工データ：テスト平均対数尤度によるモデルの性能比較. 4つの異なるスパースネスを持つデータセットに対して, 5交差検定法を行った. テストセットへの平均対数尤度の平均値と標準偏差を示す. 各実験で最も高い平均対数尤度を示し, 統計的有意差 ($p = 0.05$) を示したモデルに対応する数値を太字で表した

Table 1 Comparison on the average test log likelihood of non-zero entries among different sparseness. Means and one standard deviations of four data sets are presented.

Sparseness	NMF	VBNMF	PMF	NM2F
0%	-1.24 ± 0.21	-2.72 ± 0.03	-2.47 ± 0.239	-1.03 ± 0.09
9%	-19.39 ± 2.82	-8.49 ± 0.60	-13.00 ± 2.74	-0.99 ± 0.08
99%	-42.45 ± 6.30	-14.55 ± 1.40	-16.25 ± 6.05	-1.07 ± 0.25
99.9%	-43.25 ± 33.45	-15.30 ± 6.30	-12.85 ± 11.20	-0.86 ± 0.55

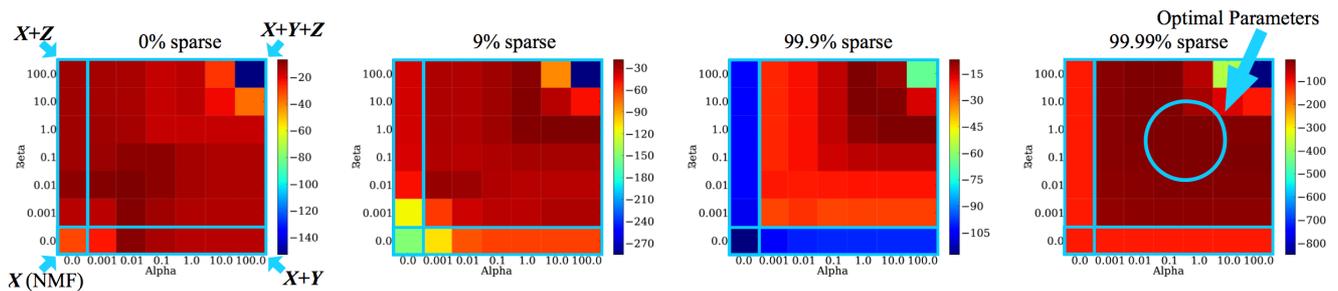


図 3 テストセットへの平均対数尤度とスケール変数 α (水平軸) と β (鉛直軸) の関係
 Fig. 3 The average test log likelihood versus the scaling parameters α (horizontal) and β (vertical).

する.

モデル性能と観測行列のスパースネスの関係を検証するため, 観測行列のスパースネス (行列の全要素に占める零要素の割合) を 0%, 9%, 99%, 99.9% の 4 段階に設定した実験を行う. なお, 補助行列 Y, Z は, 全要素が非零とする. スケール変数は交差検定法によって決定する. スケール変数 α, β の選択によるモデルの性能への影響を確認するために, 各スパースネスの設定においてスケール変数のグリッドサーチを行う.

4.3 結果

人工データを用いた実験結果を表 1 に示す. また, スケール変数のグリッドサーチ結果を図 3 に示す. NM2F は, データのスパースネスが 9% 以上の場合において, 他手法よりも統計的有意差 ($p = 0.05$) を満たす高い対数尤度を示している. データのスパースネスが高まるにつれて, 他手法との差異は大きくなっている. スケール変数に対するグリッドサーチの結果 (図 3) を確認すると, 適切なスケール変数の選択による NM2F の性能向上が確認できる. ただし, NM2F はスケール変数のオーダが一致していれば対数尤度は安定して上昇するため, スケール変数の選択に対してロバストであるといえる. 総じて補助行列を用いることによって, NM2F は他手法に比べ非零要素の推定精度が向上しているといえる.

4.4 実データ実験

実データ実験では 2 つのデータセットを用いる. なお VBNMF と PMF は計算量が大きいため実験から除外した. NMF と NM2F の汎化性能を比較するために, 観測行列 X のテストセットの平均対数尤度を用いる. これは実データがある確率構造から生成されると仮定した際, トレーニングセットから経験的に学習した NMF と NM2F のモデルが, それぞれ真の確率構造をどれだけ正確に推定できたか, という指標になる. 補助行列 Y, Z は, NM2F のみが行列分解に利用しているため汎化性能の比較には用いない.

4.4.1 Last.fm データ

Last.fm データセットは *Hetrec 2011*^{*2} から提供されているものを利用する. データセットには 1,892 人のユーザ, 17,632 組のアーティスト, 11,946 個のタグが含まれている. ユーザ, アーティスト, タグ, フレンド関係にあるユーザのインデックスを, i, j, n, m と定める. $x_{i,j}$ はあるユーザが i が, アーティスト j の楽曲を視聴した回数とする. $X \in \mathbb{R}_+^{1,892 \times 17,632}$ を観測行列とする. $y_{n,j}$ は, タグ n が, アーティスト j にタグ付けされた回数とする. $Y \in \mathbb{R}_+^{11,946 \times 17,632}$ を補助行列とする. $z_{i,m}$ は, ユーザ i とユーザ m の友人関係を表すとする. もしユーザ同士が友人関係にあれば $z_{i,m} = 1$ となり, そうでなければ $z_{i,m} = 0$ となる. $Z \in \mathbb{R}_+^{1,892 \times 1,892}$ を補助行列とする. NMF と NM2F の性能を, X のテストセットに対する平均対数尤度によっ

*2 <http://ir.ii.uam.es/hetrec2011/>

表 2 実データ：テストセットに対する平均対数尤度の平均と標準偏差

Table 2 Comparison of average test log likelihood of nonzero entries among different data sets.

Data set	NMF	VBNMF	PMF	NM2F
Last.fm	-6.90 ± 0.03	N/A	N/A	-6.17 ± 0.03
Togetter	-27.27 ± 0.23	N/A	N/A	-12.97 ± 0.48

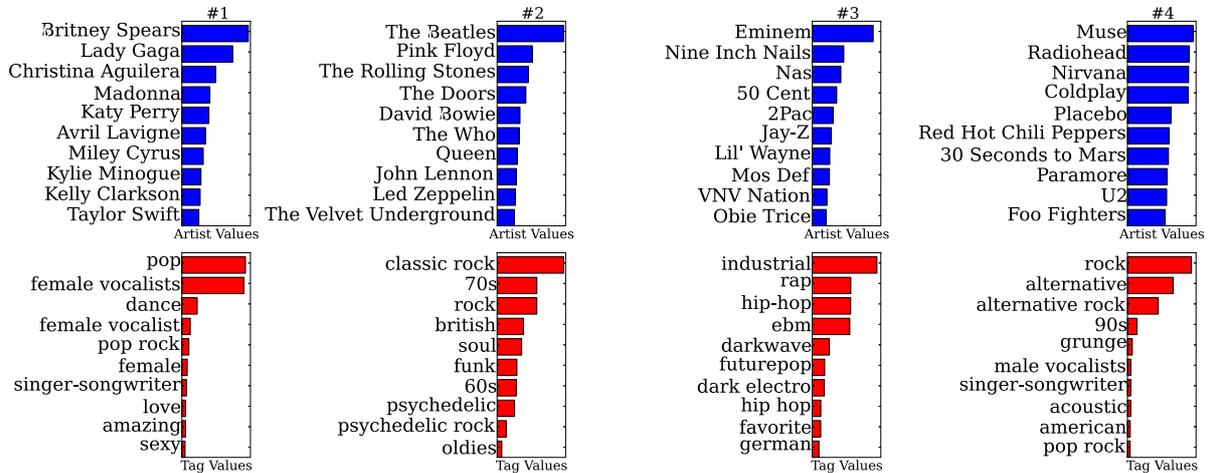


図 4 NM2F：基底で高い値をとったアーティストとタグ (w_k ：上部青色棒グラフ, a_k ：下部赤色棒グラフ)

Fig. 4 Artists and Tags with the highest values in bases of NM2F for Last.fm data.

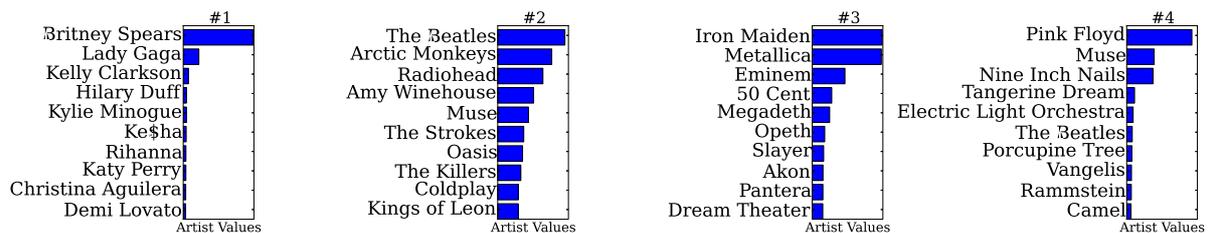


図 5 NMF：基底で高い値をとったアーティストとタグ (w_k ：青色棒グラフ)

Fig. 5 Artists with the highest values in bases of NMF for Last.fm data.

て評価する．行列 X はスパースネスが 99.75% とスパース性が高い．基底数 K は予備実験の結果から $K = 20$ とした．

4.4.2 Togetter データ

ソーシャルキュレーションサービスである Togetter データ (収集期間は 2009 年 9 月から 2012 年 7 月まで) を利用する．Togetter は、ユーザが Twitter 上の発言を集めて作成する、まとめ記事を提供するサービスである．まとめ記事におけるユーザの発言頻度を bag of words [3] として表現した観測行列を、 $X \in \mathbb{R}_+^{1,823,184 \times 235,086}$ とする．次元はユーザに対応し、データ点はまとめ記事に対応する． X の要素 $x_{i,j}$ は非負値であり、スパースネスが 99.9982% のスパースな行列となる．行列 X は約 180 万 \times 20 万要素の大きさとなるが、NMF と NM2F の乗法的更新則はデータのスパース性を利用し効率的に計算を行える．まとめ記事に対する補助データとして、まとめ記事での単語頻度を bag of words として表現した、補助行列を $Y \in \mathbb{R}_+^{165,046 \times 235,086}$

とする．次元は語彙種類に対応し、データ点はまとめ記事に対応する． Y は、スパースネスが 98.77% のスパースな行列となる．ユーザごとの補助データとして、Twitter でのフォロー数とリスト数 \tilde{z} を、 $z = \log(\tilde{z} + 1)$ と変換した数値を、補助行列 $Z \in \mathbb{R}_+^{1,823,184 \times 2}$ とする．基底数 K は予備実験の結果から $K = 200$ とした．

4.5 結果

実データ実験の定量評価の結果を表 2 に示す．2 つの実データにおいて NM2F は NMF に対して優位な数値を示している．表 2 から NM2F の NMF に対する優位性を確認できた．

次に NM2F の性能を定性的に評価する．NM2F によって得られた基底 W, A を図 4, 図 6 に示す．NMF によって得られた基底 W を図 5 に示す．それぞれの図には 4 つの基底を示しており、各基底で高い値をとった上位の要素を示した．

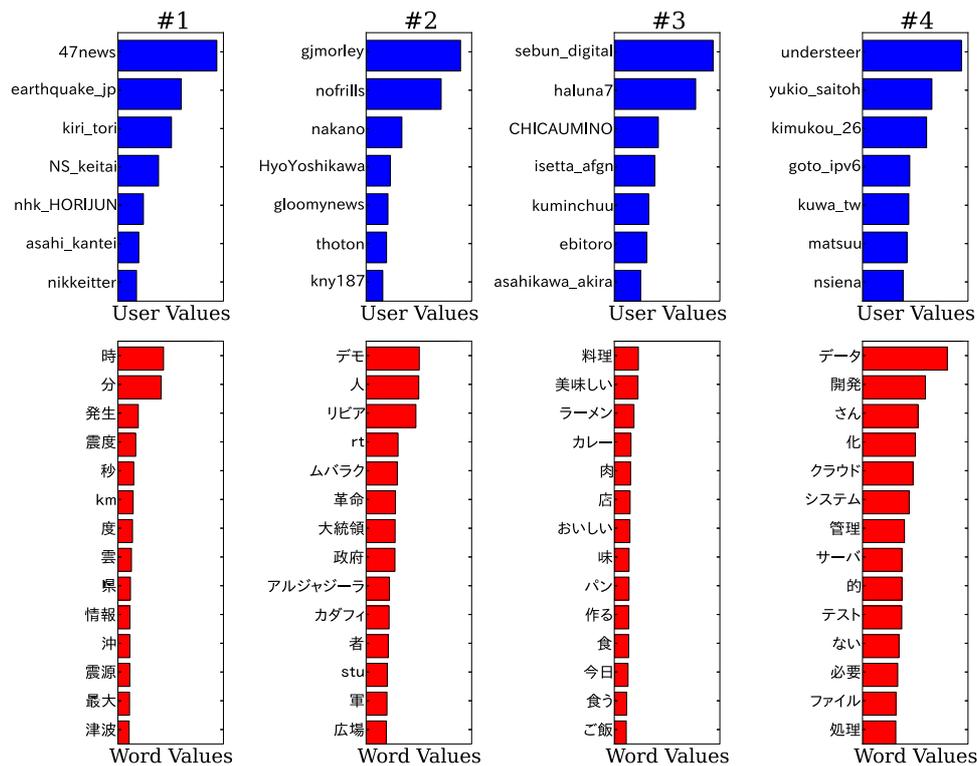


図 6 NM2F：基底で高い値をとった Twitter ユーザと単語 (w_k ：上部青色棒グラフ, a_k ：下部赤色棒グラフ)

Fig. 6 Users and Words with the highest values in bases of NM2F for social curation data.

Last.fm データセットの解析結果を検証する。図 4 の基底#1 の W では、*Britney Spears* や *Lady Gaga* などの女性歌手が高い値を持つ。 A では、pop と female vocalist のタグが高い値を持つ。NM2F によって女性ポップ歌手とそれに対応するタグが抽出されたといえる。同様に、基底#2 には *The Beatles* や *Pink Floyd*, *The Rolling Stones* などのロック歌手が高い値を持ち、classic rock や 70s などのタグが同時に高い値を持つ。基底#3 で高い値を持つ歌手は、*Eminem* や *Nine Inch Neils* などのヒップホップ歌手があり、高い値を持つタグには industrial, rap, hip-hop などがある。最後に基底#4 では、*Muse*, *Radiohead*, *Nirvana* などのロック歌手と、rock, alternative, alternative rock などのタグが基底が現れた。いずれの基底においても歌手のジャンルとジャンルに関係性の高いタグが同時に抽出されている。NM2F はユーザの視聴履歴、アーティストのタグ情報、ユーザの友人関係を活用することで、理解しやすい基底の学習を行っていることが分かる。

上の結果と NMF の分解結果 W を結果を定性的に比較する。図 5 には、NMF から得られた基底のうち、図 4 の基底で最も高い値を持った歌手が含まれているものを示した。基底#1 では、図 4 の基底#1 のほぼ同一の歌手が抽出された。しかし、基底#2 では、NM2F では別々の基底に分類された *The Beatles* と *Muse* が同じ基底で高い値を持つ。基底#3 では、*Iron Meiden* や *Metalica* など異なる

ジャンルの歌手が高い値を持つ。最後に基底#4 では、基底#2 の場合と同様に *Pink Floyd*, *Muse*, *Nine Inch Nails* が同一の基底に現れた。以上から、NM2F はユーザ視聴履歴だけでは抽出できなかった精度での行列分解を実現した結果、高いテストセットの平均対数尤度を達成したと考えられる。

Togetter データセットの解析結果を検証する。図 6 の基底#1 には、*47news* や *earthquake.jp*, *NS_Keitai* などのニュース速報や地震速報を発信する Twitter アカウントやマスコミの公式アカウントが高い値を持ち、「時、分、発生、震度、秒」などの単語が高い値を持つ。地震や津波などの災害情報を発信した際に注目を集めたアカウントが抽出されたと考えられる。基底#2 には、プロフィールにジャーナリズムやニュースなどをへの関心を書いた Twitter アカウントが高い値を持ち、「デモ、人、リビア、ムバラク、革命」などの単語が高い値を持つ。アラブの春に関連した情報発信を行ったアカウントが抽出されたと考えられる。家庭雑誌の広報アカウントである *seibun_digital* や芸能人、作家のアカウントが高い値を持ち、「料理、美味しい、ラーメン、カレー」などの単語が高い値を持つ。食生活に関するアカウントが抽出されたと考えられる。最後に、基底#4 には、ウェブサービスやクラウドに関するアカウントが抽出されており、「データ、開発、クラウド、システム」などの単語が高い値を持つ。以上から、ユーザの共起頻度のみ

だけでなく、単語の共起頻度や Twitter 上の人気度などの複合的な行列を同時に分解することで、基底を多角的に解釈が可能になり、かつ非負制約を課すことにより、理解しやすい基底が抽出可能になるといえる。

5. 考察

本稿では、高度にスパースな観測行列を分解する際に、他の補助行列を用いて、非負制約下で行列を同時分解する手法として、複合非負値行列因子分解 (Non-negative Multiple Matrix Factorization: NM2F) を提案した。NM2F で扱う問題を定式化したのち、一般化 KL ダイバージェンスを用いて再構成誤差を定め、補助関数法によるパラメータ推定法を提示した。NM2F のブロック未定義領域あり NMF としての解釈と、ポアソン分布を用いた確率的生成モデルとしての解釈を行った。人工データ実験によって、NM2F と他の既存手法の汎化性能を比較し、NM2F の優位性を示した。さらに実データ実験によって、NM2F が実データでも定量的に有効な性能を示した。NM2F の学習した基底行列 W と補助基底行列 A を照らし合わせることで、ソーシャルメディアデータの解析結果の解釈が容易になることを定性的に示した。

提案した NM2F が持つ課題あるいは拡張の余地としては、補助行列が 3 つ以上存在する場合の複合行列分解、あるいは非負制約を課さない複合行列分解の検討などがあげられる。また、非負制約が直感的に理解しやすい行列分解結果を示すとされているが、被験者実験などを用いた認知心理学的なアプローチから定量的に評価する必要がある。再構成性誤差においては、一般的な β ダイバージェンス [9] での定式化や、データごとに異なる β ダイバージェンスを用いた定式化を行い、データの生成過程をより柔軟に記述するモデルを検討する必要がある。

参考文献

- [1] Abdallah, S.A. and Plumbley, M.D.: Polyphonic music transcription by non-negative sparse coding of power spectra, *Proc. ISMIR* (2004).
- [2] Aharon, M., Elad, M. and Bruckstein, A.M.: On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them, *Linear algebra and its applications*, Vol.416, No.1, pp.48-67 (2006).
- [3] Blei, D.M., Ng, A.Y. and Jordan, M.I.: Latent Dirichlet allocation, *Journal of Machine Learning Research*, Vol.3, pp.993-1022 (2003).
- [4] Bouchard, G., Guo, S. and Yin, D.: Convex collective matrix factorization, *Proc. AISTATS* (2013).
- [5] Cai, D., He, X. and Han, J.: Graph regularized non-negative matrix factorization for data representation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.33, pp.1548-1560 (2011).
- [6] Cao, B., Shen, D., Sun, J.T., Wang, X., Yang, Q. and Chen, Z.: Detect and track latent factors with online nonnegative matrix factorization, *Proc. IJCAI* (2007).
- [7] Cemgil, A.T.: Bayesian inference for nonnegative matrix factorisation models, *Computational Intelligence and Neuroscience*, Vol.4, pp.1-17 (2009).
- [8] Cichocki, A., Zdunek, R. and Amari, S.: Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization, *Proc. ICA* (2007).
- [9] Cichocki, A., Zdunek, R., Phan, A.H. and Amari, S.: *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis*, Wiley (2009).
- [10] Dikmen, O. and Févotte, C.: Maximum marginal likelihood estimation for nonnegative dictionary learning in the Gamma-Poisson model, *IEEE Trans. Signal Processing*, Vol.60, pp.5163-5175 (2012).
- [11] Ding, C., Li, T. and Peng, W.: NMF and PLSI: Equivalence and a hybrid algorithm, *Proc. SIGIR* (2006).
- [12] Eggert, J. and Korner, E.: Sparse coding and nmf, *Proc. IJCNN* (2004).
- [13] Gupta, M.D. and Xiao, J.: Non-negative matrix factorization as a feature selection tool for maximum margin classifiers, *Proc. CVPR* (2011).
- [14] Gupta, S.K., Phung, D., Adams, B. and Venkatesh, S.: A matrix factorization framework for jointly analyzing multiple nonnegative data sources, *Text mining workshop, in conjunction with SDM* (2011).
- [15] Gupta, S.K., Phung, D., Adams, B., Tran, T. and Venkatesh, S.: Nonnegative shared subspace learning and its application to social media retrieval, *Proc. SIGKDD* (2010).
- [16] Hoffman, M.D., Blei, D.M. and Cook, P.R.: Bayesian nonparametric matrix factorization for recorded music, *Proc. ICML* (2011).
- [17] Hofmann, T.: Probabilistic latent semantic indexing, *Proc. SIGIR* (1999).
- [18] Hoyer, P.O.: Non-negative sparse coding, *Proc. NNSP* (2002).
- [19] Hoyer, P.O.: Non-negative matrix factorization with sparseness constraints, *Journal of Machine Learning Research*, Vol.5, pp.1457-1469 (2004).
- [20] Koren, Y., Bell, R. and Volinsky, C.: Matrix factorization techniques for recommender systems, *IEEE Computer*, Vol.42, No.8, pp.30-37 (2009).
- [21] Lee, D.D. and Seung, H.S.: Learning the parts of objects by non-negative matrix factorization, *Nature*, Vol.401, pp.788-791 (1999).
- [22] Lee, D.D. and Seung, H.S.: Algorithms for non-negative matrix factorization, *Proc. NIPS* (2000).
- [23] Lin, J., Snow, R. and Morgan, W.: Smoothing techniques for adaptive online language models: Topic tracking in tweet streams, *Proc. SIGKDD* (2011).
- [24] Liu, C., Yang, H., Fan, J., He, L. and Wang, Y.: Distributed nonnegative matrix factorization for web-scale dyadic data analysis on MapReduce, *Proc. WWW* (2010).
- [25] Liu, H., Yang, Z., Wu, Z. and Li, X.: A-optimal non-negative projection for image representation, *Proc. CVPR* (2012).
- [26] Ma, H., Yang, H., Lyu, M.R. and King, I.: Sorec: Social recommendation using probabilistic matrix factorization, *Proc. CIKM* (2008).
- [27] Mei, Q., Cai, D., Zhang, D. and Zhai, C.: Topic modeling with network regularization, *Proc. WWW* (2008).
- [28] Nakano, M., Roux, J.L., Kameoka, H., Nakamura, T., Ono, N. and Sagayama, S.: Bayesian nonparametric spectrogram modeling based on infinite factorial infinite hidden Markov model, *Proc. WASPAA* (2011).

[29] Noel, J., Sanner, S., Tran, K., Christen, P., Xie, L., Bonilla, E.V., Abbasnejad, E. and Penna, N.D.: New objective functions for social collaborative filtering, *Proc. WWW* (2012).

[30] Purushotham, S., Liu, Y. and Kuo, C.C.J.: Collaborative topic regression with social matrix factorization for recommendation systems, *Proc. ICML* (2012).

[31] Roux, J.L., Kameoka, H., Ono, N., Cheveigne, A.D. and Sagayama, S.: Computational auditory induction as a missing-data model-fitting problem with Bregman divergence, *Speech Communication*, Vol.53, No.5, pp.658–676 (2011).

[32] Salakhutdinov, R. and Mnih, A.: Probabilistic matrix factorization, *Proc. NIPS* (2008).

[33] Schmidt, M.N., Winther, O. and Hansen, L.K.: Bayesian non-negative matrix factorization, *Proc. ICASSP* (2009).

[34] Smaragdis, P. and Brown, J.C.: Non-negative matrix factorization for polyphonic music transcription, *Proc. WASPA* (2003).

[35] Steyvers, M., Smyth, P. and Griffiths, T.: Probabilistic author-topic models for information discovery, *Proc. SIGKDD* (2004).

[36] Wang, C. and Blei, D.M.: Collaborative topic modeling for recommending scientific articles, *Proc. SIGKDD* (2011).

[37] Wang, F., Tan, C., König, A.C. and Li, P.: Efficient document clustering via online nonnegative matrix factorizations, *Proc. SIAM* (2011).

[38] Xu, W., Liu, X. and Gong, Y.: Document clustering based on non-negative matrix factorization, *Proc. SIGIR* (2003).

付 録

A.1 補助関数法

補助関数法による NMF の乗法的更新則導出法を述べる。再構成誤差に一般化 KL ダイバージェンス (*gKL*) を用いた NMF の再構成誤差は、

$$\begin{aligned} \mathcal{D}(\mathbf{X}|\mathbf{W}, \mathbf{H}) &= \sum_{i=1}^I \sum_{j=1}^J \left(\hat{x}_{i,j} - x_{i,j} \log \sum_{k=1}^K w_{i,k} h_{k,j} \right) + \text{const}, \end{aligned} \quad (\text{A.1})$$

である。式 (A.1) を最小化する乗法的更新則を補助関数法によって求める。まず補助変数 $r_{i,j,k} \geq 0$, $(1, \dots, K)$ を導入し、Jensen の不等式によって補助関数 (上限) を求める。式 (A.1) に補助変数を挿入し、

$$\begin{aligned} &\sum_{i=1}^I \sum_{j=1}^J \left(\hat{x}_{i,j} - x_{i,j} \log \sum_{k=1}^K w_{i,k} h_{k,j} \right) \\ &= \sum_{i=1}^I \sum_{j=1}^J \left(\hat{x}_{i,j} - x_{i,j} \log \sum_{k=1}^K r_{i,j,k} \frac{w_{i,k} h_{k,j}}{r_{i,j,k}} \right), \end{aligned} \quad (\text{A.2})$$

と変形する。Jensen の不等式を用いて、再構成誤差 \mathcal{D} に対する補助関数 (上限) $\mathcal{F}(\mathbf{X}|\mathbf{W}, \mathbf{H}, r_{i,j,k})$ を求める。

$$\begin{aligned} \mathcal{D}(\mathbf{X}|\mathbf{W}, \mathbf{H}) &= \sum_{i=1}^I \sum_{j=1}^J \left(\hat{x}_{i,j} - x_{i,j} \log \sum_{k=1}^K r_{i,j,k} \frac{w_{i,k} h_{k,j}}{r_{i,j,k}} \right) \\ &\leq \sum_{i=1}^I \sum_{j=1}^J \left(\hat{x}_{i,j} - x_{i,j} \sum_{k=1}^K r_{i,j,k} \log \frac{w_{i,k} h_{k,j}}{r_{i,j,k}} \right) \\ &\triangleq \mathcal{F}(\mathbf{X}|\mathbf{W}, \mathbf{H}, r_{i,j,k}). \end{aligned} \quad (\text{A.3})$$

ここでは例として $w_{i,k}$ について再構成誤差を最小化する。補助関数を $w_{i,k}$ について偏微分すると、

$$\frac{\partial \mathcal{F}}{\partial w_{ik}} = \sum_{j=1}^J \left(h_{k,j} - x_{i,j} \frac{r_{i,j,k}}{w_{i,k}} \right), \quad (\text{A.4})$$

を得る。 $\frac{\partial^2 \mathcal{F}}{\partial^2 w_{ik}} \geq 0$ であるから、補助関数は $w_{i,k}$ に関して凹である。すなわち停留点が $w_{i,k}$ に関して最小値を与える。 $\frac{\partial \mathcal{F}}{\partial w_{ik}} = 0$ を解くと、

$$w_{i,k}^{\text{new}} = \frac{\sum_{j=1}^J x_{i,j} r_{i,j,k}}{\sum_{j=1}^J h_{k,j}} \in \mathbf{W}^{\text{new}}, \quad (\text{A.5})$$

を得る。このとき、

$$\mathcal{D}(\mathbf{X}|\mathbf{W}^{\text{new}}, \mathbf{H}) \leq \mathcal{F}(\mathbf{X}|\mathbf{W}^{\text{new}}, \mathbf{H}, r_{i,j,k}), \quad (\text{A.6})$$

である。Jensen の不等式の性質から、式 (A.6) の等号成立条件は、

$$r_{i,j,k} = \frac{w_{i,k} h_{k,j}}{\sum_{k=1}^K w_{i,k} h_{k,j}}, \quad (\text{A.7})$$

である。したがって補助変数を上記の値に設定することで、補助関数最小化問題の解は元の再構成誤差最小化問題の解と一致する。以上から、式 (A.7) を式 (A.5) に代入することで、乗法的更新則、

$$w_{i,k}^{\text{new}} = w_{i,k} \frac{\sum_{j=1}^J \frac{x_{i,j}}{\hat{x}_{i,j}} h_{k,j}}{\sum_{j=1}^J h_{k,j}}, \quad (\text{A.8})$$

が得られる。 $h_{k,j}$ に関する乗法的更新則も同様に導出される。

A.2 補助関数法による NM2F の乗法的更新則導出

式 (7) を最小化しよう各変数の乗法的更新則を導出する。 $h_{k,j}$, $a_{n,k}$, $b_{k,m}$ についても以下と同様に乗法的更新則は求められる。

$$\begin{aligned}
 & \mathcal{D}(\mathbf{X}, \mathbf{Y}, \mathbf{Z} | \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}; \alpha, \beta) \\
 & \simeq \sum_{i=1}^I \sum_{j=1}^J \left(\hat{x}_{i,j} - x_{i,j} \log \sum_{k=1}^K w_{i,k} h_{k,j} \right) \\
 & + \alpha \sum_{n=1}^N \sum_{j=1}^J \left(\hat{y}_{n,j} - y_{n,j} \log \sum_{k=1}^K a_{n,k} h_{k,j} \right) \\
 & + \beta \sum_{i=1}^I \sum_{m=1}^M \left(\hat{z}_{i,m} - z_{i,m} \log \sum_{k=1}^K w_{i,k} b_{k,m} \right). \quad (\text{A.9})
 \end{aligned}$$

補助変数として, $r_{i,j,k}$, $s_{n,j,k}$, $t_{i,m,k}$ ($\sum_k r_{i,j,k} = 1$, $\sum_k s_{n,j,k} = 1$, $\sum_k t_{i,m,k} = 1$) を定める. \mathcal{D} の上限 \mathcal{F} は Jensen の不等式を用いて,

$$\begin{aligned}
 & \mathcal{D}(\mathbf{X}, \mathbf{Y}, \mathbf{Z} | \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}; \alpha, \beta) \\
 & \leq \sum_{i=1}^I \sum_{j=1}^J \left(\hat{x}_{i,j} - x_{i,j} \sum_{k=1}^K r_{i,j,k} \log \frac{w_{i,k} h_{k,j}}{r_{i,j,k}} \right) \\
 & + \alpha \sum_{n=1}^N \sum_{j=1}^J \left(\hat{y}_{n,j} - y_{n,j} \sum_{k=1}^K s_{n,j,k} \log \frac{a_{n,k} h_{k,j}}{s_{n,j,k}} \right) \\
 & + \beta \sum_{i=1}^I \sum_{m=1}^M \left(\hat{z}_{i,m} - z_{i,m} \sum_{k=1}^K t_{i,m,k} \log \frac{w_{i,k} b_{k,m}}{t_{i,m,k}} \right) \\
 & \triangleq \mathcal{F}(\mathbf{X}, \mathbf{Y}, \mathbf{Z} | \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}; \alpha, \beta, r_{i,j,k}, s_{n,j,k}, t_{i,m,k}), \quad (\text{A.10})
 \end{aligned}$$

と求められる. 等号は,

$$\begin{aligned}
 r_{i,j,k} &= \frac{w_{i,k} h_{k,j}}{\sum_{k=1}^K w_{i,k} h_{k,j}}, \quad s_{n,j,k} = \frac{a_{n,k} h_{k,j}}{\sum_{k=1}^K a_{n,k} h_{k,j}}, \\
 t_{i,m,k} &= \frac{w_{i,k} b_{k,m}}{\sum_{k=1}^K w_{i,k} b_{k,m}}. \quad (\text{A.11})
 \end{aligned}$$

のとき成立する. $w_{i,k}$ について \mathcal{F} を偏微分すると,

$$\begin{aligned}
 & \frac{\partial \mathcal{F}}{\partial w_{i,k}} \\
 & = \sum_{j=1}^J \left(h_{k,j} - x_{i,j} \frac{r_{i,j,k}}{w_{i,k}} \right) + \beta \sum_{m=1}^M \left(b_{k,m} - z_{i,m} \frac{t_{i,m,k}}{w_{i,k}} \right), \quad (\text{A.12})
 \end{aligned}$$

を得る. $\frac{\partial \mathcal{F}}{\partial w_{i,k}} = 0$ とおくと, 式 (A.12) は以下のように書き下せる.

$$w_{i,k} = \frac{\sum_{j=1}^J x_{i,j} r_{i,j,k} + \beta \sum_{m=1}^M z_{i,m} t_{i,m,k}}{\sum_{j=1}^J h_{k,j} + \beta \sum_{m=1}^M b_{k,m}}. \quad (\text{A.13})$$

以上から, $w_{i,k}$ の乗法的更新則は,

$$w_{i,k}^{\text{new}} = w_{i,k} \frac{\sum_{j=1}^J \frac{x_{i,j}}{\hat{x}_{i,j}} h_{k,j} + \beta \sum_{m=1}^M \frac{z_{i,m}}{\hat{z}_{i,m}} b_{k,m}}{\sum_{j=1}^J h_{k,j} + \beta \sum_{m=1}^M b_{k,m}}, \quad (\text{A.14})$$

となる. NMF も含めて, 乗法的更新則では観測値と推定パラメータの要素積を計算するが, 観測値が零の場合は計

算を省略, 効率化できる. 乗法的更新則の計算を効率的に行うために X の第 j 列ベクトルにおける非零要素の集合を \mathcal{I}_j とし, 第 j 列ベクトルの和を $\sum_{i \in \mathcal{I}_j}$ と定める. 同様に, X の第 i 行ベクトル, Y の第 j 列ベクトル, Z の第 i 列ベクトルの非零要素の集合を \mathcal{J}_i , \mathcal{M}_i , \mathcal{N}_j とすると, 上記の乗法的更新則は,

$$w_{i,k}^{\text{new}} = w_{i,k} \frac{\sum_{j \in \mathcal{J}_i} \frac{x_{i,j}}{\hat{x}_{i,j}} h_{k,j} + \beta \sum_{m \in \mathcal{M}_i} \frac{z_{i,m}}{\hat{z}_{i,m}} b_{k,m}}{\sum_j h_{k,j} + \beta \sum_m b_{k,m}}, \quad (\text{A.15})$$

となり, 観測行列と補助行列の非零要素の演算を効率的に行える. したがって行列のスパース性の高いほど計算速度が向上する.

なお \mathcal{F} の 2 階偏微分は,

$$\frac{\partial^2 \mathcal{F}}{\partial w_{i,k}^2} = \sum_{j=1}^J \left(x_{i,j} \frac{r_{i,j,k}}{w_{i,k}^2} \right) + \beta \sum_{j=1}^J \left(z_{i,m} \frac{t_{i,m,k}}{w_{i,k}^2} \right) \geq 0, \quad (\text{A.16})$$

となるため, $w_{i,k}$ について \mathcal{F} は convex であり, 解は局所解に向かって単調収束することが保証される. 乗法的更新則では, ALS のように解の非負制約が乗法的更新則で明示的に課されていないが, 変数を非負値で初期化すれば, おのずと非負制約の下で解が求まる.



竹内 孝

2009 年早稲田大学理工学部電気・情報生命工学科卒業. 2011 年同大学大学院修士課程修了. 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機械学習の研究に従事.



石黒 勝彦 (正会員)

2004 年東京大学工学部機械情報工学科卒業. 2006 年同大学大学院情報理工系研究科知能機械情報学専攻修士課程修了. 同年日本電信電話株式会社 (NTT) 入社. 現在, NTT コミュニケーション科学基礎研究所研究員. 博士 (工学). 統計的機械学習, 時系列データ解析, 関係データ解析の研究に従事. 電子情報通信学会, IEEE 各会員.



木村 昭悟

2000年東京工業大学大学院理工学研究科電気・電子工学専攻修士課程修了。同年日本電信電話株式会社 (NTT) 入社。2007年東京工業大学大学院理工学研究科集積システム専攻博士課程修了。博士 (工学)。現在, NTT コミュニケーション科学基礎研究所メディア情報研究部メディア認識研究グループ主任研究員。実世界・ウェブ世界での人間の行動を手掛かりに, 画像・映像を含むマルチメディアコンテンツの内容を認識・理解する研究と, その基礎技術の開発に従事。電子情報通信学会, IEEE, ACM SIGMM/SIGKDD 各会員。



澤田 宏 (正会員)

1991年京都大学工学部情報工学科卒業。1993年同修士課程修了。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 およびアーキテクチャの研究に従事。2000年より, 信号処理, 特に独立成分分析を用いたブラインド音源分離の研究に従事。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。