

非負制約下における複合行列分解

竹内 孝^{1,a)} 石黒 勝彦¹ 木村 昭悟¹ 澤田 宏¹

概要: ユビキタスコンピューティングやセンサーネットワークの普及に伴い、多種多様なデータが取得されるようになった。これらの大規模かつ複合的なデータを、各モダリティに対するドメイン知識を活用し特徴量抽出しつつ、包括的に解析できれば、データの多角的な解釈が可能になると考えられる。本稿では、複合的なデータを解析するための統計的機械学習のアプローチとして、複数のデータを同時解析する枠組みである、Non-negative Multiple Matrix Factorization (NMMF) を提案する。NMMF は観測行列の分解時に2つの補助行列を利用する手法である。本稿では、NMMF を Non-negative Matrix Factorization (NMF) の一般化として定式化し、実験によって、NMMF の定量的な性能を評価する。

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

KOH TAKEUCHI^{1,a)} KATSUHIKO ISHIGURO¹ AKISATO KIMURA¹ HIROSHI SAWADA¹

Abstract: Analyzing highly sparse data is often falling into cheap results. To compensate data sparseness, in this paper, we introduce a novel machine learning technique named Non-negative Multiple Matrix Factorization (NMMF). NMMF factorizes multiple matrices simultaneously under a non-negative constraint. Therefore NMMF can extract intuitively interpretable bases and coefficients from multiple matrices. We formulate NMMF 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 NMMF and other existing techniques in both the quantitative and qualitative ways. NMMF shows better performance than other techniques on both synthetic and the real-world data set.

Keywords: Machine Learning, Data Mining, Matrix Factorization

1. はじめに

Non-negative Matrix Factorization (NMF) [10] は行列分解手法の1つである。その応用範囲は広く、音声音響信号処理 [16], 画像処理 [8], [10], テキストマイニング [19], Brain Computer Interface [5] など多岐に渡る。NMF による行列分解は、しばしばデータから直感的に理解しやすい結果になると報告されている。これは NMF の非負制約が、行列分解結果にスパース制約として働くことに起因している。非負制約に加えて、他のスパース制約を課した NMF [4], [8] も提案されている。しかし、一般的に NMF を含む行列分

解手法は、観測行列が高度に疎な場合には、汎化誤差の大きい、望ましくない解析結果を示すことが知られている [1]。

近年、人間活動に伴う履歴データや環境センシングデータなど、多種多様な大規模データが観測されている。これらのデータを解析することが出来れば、解析対象の事象に対して、多角的な解釈が可能になると考えられる。しかし、これらのデータはしばしば高次元であるために、高度に疎な場合が多く、データ解析は困難である [2], [9], [11]。

この問題を解決するための1つの方策として、複数のデータからなる、複合データの同時解析があげられる。例えば、多種類のセンサーによって、ある事象が音声、映像、位置情報、テキストなど、多種類のデータとして観測される状況を仮定しよう。このとき、観測されるデータは同一の事象

¹ NTT Communication Science Laboratories, Kyoto 619-0237, Japan

^{a)} takeuchi.koh@lab.ntt.co.jp

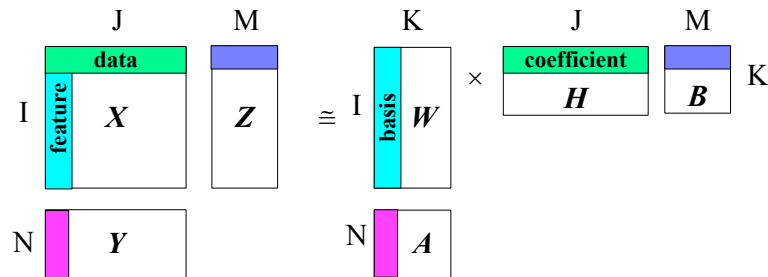


図 1 観測行列 X を基底行列 W と係数行列 H に分解する。その際、補助行列 Y と Z を同時に分解する。

を、別々のモーダルにおいて観測したデータであるため、それぞれは事象に起因した関係性を持つと考えられる。このような関係をもつ別のデータを補助的な情報として用いることが出来れば、単一データを解析した場合と比べて、解析精度が向上すると予想される。

このアイデアは、協調フィルタリング分野におけるアイテム推薦などで、既に有効性が実証されている。協調フィルタリングの分野では、もともと行列分解法が広く用いられている [9]。幾つかの研究 [13], [17] では、行列の疎性を解決するために、解析対象の観測行列に関する補助的な観測行列を援用することで、行列分解時に利用可能な情報量を増加させる方策を取っている。これらの手法はアイテム推薦の精度を向上させている。しかし、非負制約の場合は現時点では検討されていない。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り、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、データ行列の分解精度を、補助的な情報を併用することで向上する手法を検討する。解析対象のデータと補助的なデータを用いた非負制約下の行列分解手法として、Non-negative Multiple Matrix Factorization (NMMF) を提案する。NMMF は補助的なデータを補助行列として扱う。ここでは補助行列とは、観測行列と行あるいは列を共有するものとする (図 1)。NMMF によって全ての行列を同時に分解することで、高度に疎な観測行列の分解における、分解精度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、NMMF 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、NMMF 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、NMMF を確率モデルとして解釈すると同時に、ブロック状の未定義領域をもつ NMF としても解釈する。人工データと実世界データの実験によって、NMMF は高度に疎な行列の分解において、定量的に良好な性能を持つことを示す。

本稿の構成を述べる。第 2 節では、本稿に関連する研究を確認する。第 3 節では、提案手法である NMMF を提案する。第 4 節では、NMMF と他の既存手法の性能を、人工データと実データの実験を行い比較する。第 5 節では、本稿の結論を述べる。

2. 関連研究

Non-negative Matrix Factorization (NMF)[10] は、教師無し行列分解手法であり、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、近似誤差を最小化するように定式化されている。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いこと、非負制約によってスパースな解を得られるという好ましい特徴をもつ。

NMF で利用される距離基準にはユークリッド距離や一般化 KL 距離、板倉斉藤距離などがある [5]。一般化 KL 距離を用いた NMF は、Probabilistic Latent Semantic Analysis (pLSA)[7] と等価であることが示されており [6]、Latent Dirichlet Allocation (LDA)[2] を代表とするトピックモデルとしても利用できることが知られている。しかし、これらの研究は、単一の行列を分解することに焦点を当てており、複数の行列を同時に分解する論文は我々の知る限り提案されていない。

また NMF のパラメータ推定は、分散並列化によって高速化することが可能であり [12]、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 [3], [18]。本稿では、NMMF 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、NMMF はスケールアップする。また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。

行列分解手法の拡張は、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) を拡張したものである。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 [17] 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 [13], [14] が存在する。いずれの場合も補助データを用いたことによる推薦精度の改善が示唆されている。本稿で提案する NMMF は、これらの研究と高い関連性をもつが、研究対象のドメインや目的が異なっている。また、第 1 節で説明したように、上記のモデルは非負制約に関する検討を行っていない。

3. Non-negative Multi Matrix Factorization

3.1 問題定式化

本稿で扱うデータの定式化を行った後、問題を述べる。本稿では、ベクトルを太字、行列を大文字の太字、転置を T で表す。

本稿では、1 つの観測データと 2 つの補助データを扱う。観測データは $i \in \{1, \dots, I\}$ の次元からなり、データ点数を $j \in \{1, \dots, J\}$ とする。 j 点目のデータにおける、 i 次元の観測値を $x_{i,j}$ とおき、 j 点目のデータの観測を、 $\mathbf{x}_j = (x_{1,j}, \dots, x_{I,j})^T \in \mathbb{R}_+^I$ とする。観測データは、 \mathbf{x}_j を列ベクトルにもつ観測行列 \mathbf{X} として、 $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_J) \in \mathbb{R}_+^{I \times J}$ 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 $n \in \{1, \dots, N\}$ とする。 j 点目の補助データにおける、 n 次元の観測値を $y_{n,j}$ とおき、 j 点目の補助データの観測を、 $\mathbf{y}_j = (y_{1,j}, \dots, y_{N,j})^T \in \mathbb{R}_+^N$ とする。補助データは、 \mathbf{y}_j を列ベクトルにもつ補助行列 \mathbf{Y} として、 $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_J) \in \mathbb{R}_+^{N \times J}$ 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 $m \in \{1, \dots, M\}$ 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、 $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_M) \in \mathbb{R}_+^{I \times M}$ と定める。行列 \mathbf{X} , \mathbf{Y} , \mathbf{Z} の概略を図 1 の左辺に示す。

本稿の目的は、観測行列 \mathbf{X} から、基底行列と係数行列を求めるものとする。この際、補助行列 \mathbf{Y} と \mathbf{Z} を用いる事ができるものとする。我々の提案手法は補助行列 \mathbf{Y} と \mathbf{Z} を明示的に利用できるため、解析精度の向上が期待できる。行列分解で補助行列の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抜き出せる利点を持つ。

観測行列と補助行列の具体例を説明する。第 4 節では、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、ある利用者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として扱う。

3.2 提案手法

NMMF は非負制約下で観測行列 \mathbf{X} と補助行列 \mathbf{Y} , \mathbf{Z} を基底行列を係数行列に分解する。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。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。NMMF 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。

NMMF では、以下の仮定をおく。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。換言すると、 \mathbf{Y} は、 \mathbf{X} とは異なる係数が、射影 \mathbf{W} によって構成される行列、 \mathbf{Z} は、 \mathbf{X} とは同一の係数 \mathbf{H} が、異なる射影によって構成される行列となる。

ただし、補助行列 \mathbf{Y} と \mathbf{Z} は、観測行列とスケールが異なるため、NMMF は、スケールの違いをスケール変数 α , β によって補正する。

観測行列は K 個の基底からなると仮定する。観測行列 \mathbf{X} の k 番目の基底ベクトルを $\mathbf{w}_k = (w_{1,k}, \dots, w_{I,k}) \in \mathbb{R}_+^I$ 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 $\mathbf{a}_k = (a_{1,k}, \dots, a_{N,k}) \in \mathbb{R}_+^N$ 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 $\mathbf{h}_j = (h_{1,j}, \dots, h_{K,j}) \in \mathbb{R}_+^K$ 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トル $\mathbf{b}_m = (b_{1,m}, \dots, b_{K,m}) \in \mathbb{R}_+^K$ 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。

NMMF は、近似誤差 $\mathcal{D}(\mathbf{X}, \mathbf{Y}, \mathbf{Z} | \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}; \alpha, \beta)$ を最小化する。

$$\begin{aligned} & \underset{\mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}}{\operatorname{argmin}} \mathcal{D}(\mathbf{X}, \mathbf{Y}, \mathbf{Z} | \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}; \alpha, \beta) \\ & \text{s.t. } \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B} \geq 0. \end{aligned} \quad (1)$$

このとき、 α , β は $0 \leq \alpha$, $0 \leq \beta$ を満たすスケール変数とした。行列 \mathbf{X} , \mathbf{Y} , \mathbf{Z} を紐付ける仮定として、 $x_{i,j}$ と $y_{n,j}$ は係数行列 \mathbf{H} を共有し、 $x_{i,j}$ と $z_{i,m}$ は基底行列 \mathbf{W} を共有するものとする。 $x_{i,j}$, $y_{n,j}$, $z_{i,m}$ の推定値を、基底と係数の重み付け線形和として、

$$\hat{x}_{i,j} = \sum_{k=1}^K w_{i,k} h_{k,j}, \quad \hat{y}_{n,j} = \sum_{k=1}^K a_{n,k} h_{k,j}, \quad \hat{z}_{i,m} = \sum_{k=1}^K w_{i,k} b_{k,m}, \quad (2)$$

と定める. すると近似誤差 \mathcal{D} は, 次のように,

$$\begin{aligned} \mathcal{D}(\mathbf{X}, \mathbf{Y}, \mathbf{Z} | \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}; \alpha, \beta) \\ &= \mathcal{D}(\mathbf{X} | \mathbf{W}, \mathbf{H}) + \alpha \mathcal{D}(\mathbf{Y} | \mathbf{A}, \mathbf{H}) + \beta \mathcal{D}(\mathbf{Z} | \mathbf{W}, \mathbf{B}) \\ &= \sum_{i=1}^I \sum_{j=1}^J d(x_{i,j} | \hat{x}_{i,j}) + \alpha \sum_{n=1}^N \sum_{j=1}^J d(y_{n,j} | \hat{y}_{n,j}) \\ &\quad + \beta \sum_{i=1}^I \sum_{m=1}^M d(z_{i,m} | \hat{z}_{i,m}), \end{aligned} \quad (3)$$

と書き下せる. このとき, $d(\cdot)$ は行列の各要素ごとの距離である. 本稿では, 距離基準に一般化 KL 距離 (the generalized Kullback-Leibler divergence: gKL) を用いる.

$$d_{gKL}(x_{i,j} | \hat{x}_{i,j}) = x_{i,j} \log \frac{x_{i,j}}{\hat{x}_{i,j}} - x_{i,j} + \hat{x}_{i,j}. \quad (4)$$

一般化 KL 距離は, NMF に広く用いられる距離基準である. ある変数 $\lambda \geq 0$ を定めると, 一般化 KL 距離は, $\lambda d_{gKL}(x_{i,j} | \hat{x}_{i,j}) = d_{gKL}(\lambda x_{i,j} | \lambda \hat{x}_{i,j})$ が成り立つ. なお, NMMF は一般化 KL 距離に限らず, ユークリッド距離や板倉齋藤距離を用いて定式化することも可能である.

3.3 NMF との比較

NMF は近似誤差 $\mathcal{D}(\mathbf{X} | \mathbf{W}, \mathbf{H})$ を最小化するように, 観測行列 \mathbf{X} から, K 個の基底からなる基底行列 \mathbf{W} と, J 個の係数ベクトルからなる係数行列 \mathbf{H} を推定する手法である.

$$\operatorname{argmin}_{\mathbf{W}, \mathbf{H}} \mathcal{D}(\mathbf{X} | \mathbf{W}, \mathbf{H}) \text{ s.t. } \mathbf{W}, \mathbf{H} \geq 0. \quad (5)$$

NMMF において, スケーリング変数 $\alpha = \beta = 0$ とするとき, NMMF の近似誤差 $\mathcal{D}(\mathbf{X}, \mathbf{Y}, \mathbf{Z} | \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}; \alpha, \beta)$ は,

$$\mathcal{D}(\mathbf{X}, \mathbf{Y}, \mathbf{Z} | \mathbf{W}, \mathbf{H}, \mathbf{A}, \mathbf{B}; \alpha, \beta) = \mathcal{D}(\mathbf{X} | \mathbf{W}, \mathbf{H}), \quad (6)$$

となる. したがって, NMMF は NMF を一般化したものと言える.

3.4 モデル推定

次の更新式によって W, H, A, B を学習し, \mathcal{D} を最小化する. 導出方法の詳細は, 付録に示す.

$$\begin{aligned} w_{i,k}^{new} &= w_{i,k} \frac{\left(\sum_{j \in \mathcal{I}_i} \frac{x_{i,j}}{\hat{x}_{i,j}} h_{k,j} + \beta \sum_{m \in \mathcal{M}_i} \frac{z_{i,m}}{\hat{z}_{i,m}} b_{k,m} \right)}{\sum_j h_{k,j} + \beta \sum_m b_{k,m}}, \\ h_{k,j}^{new} &= h_{k,j} \frac{\left(\sum_{i \in \mathcal{I}_j} \frac{x_{i,j}}{\hat{x}_{i,j}} w_{k,j} + \alpha \sum_{n \in \mathcal{N}_j} \frac{y_{n,j}}{\hat{y}_{n,j}} a_{n,k} \right)}{\sum_i w_{k,j} + \alpha \sum_n a_{n,k}}, \\ a_{n,k}^{new} &= a_{n,k} \frac{\sum_{j \in \mathcal{J}_n} \frac{y_{n,j}}{\hat{y}_{n,j}} h_{k,j}}{\sum_j h_{k,j}}, \\ b_{k,m}^{new} &= b_{k,m} \frac{\sum_{i \in \mathcal{I}_m} \frac{z_{i,m}}{\hat{z}_{i,m}} w_{i,k}}{\sum_i w_{i,k}}. \end{aligned} \quad (7)$$

更新式においても, $\alpha = 0, \beta = 0$ のとき, NMMF は NMF と一致することから, NMMF は NMF を一般化したモデルと言える. スケーリング変数 α, β はクロスバリデーションによって決定する.

3.5 確率的生成モデルとの解釈

NMMF 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 距離を利用した NMMF 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. 観測データがポアソン分布, $p(\eta | \mu) = \frac{\mu^\eta}{\eta!} \exp(-\mu)$ にしたがうとすると, モデルの対数尤度は,

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

となる. したがって, (4) 式の最小化は, 上記の確率的生成モデルの対数尤度最大化と等価である.

NMMF 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にすることで, $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ が行列 \mathbf{W} , \mathbf{W} を共有する構造が確認できる.

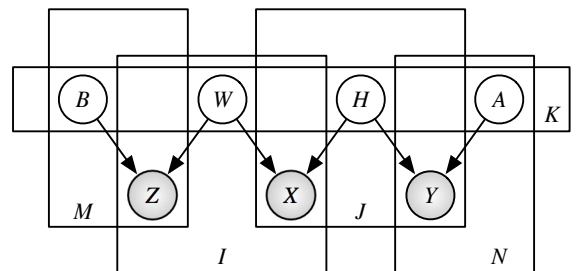


図 2 NMMF のグラフィカルモデル.

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、対数尤度を用いる。トレーニングデータセットへの対数尤度を最小化するように各パラメータを学習した後、テストデータセットへの対数尤度の平均値を求め、各モデルのデータへの性能を評価する。対数尤度の平均値は、 $\frac{1}{M} \sum_{m=1}^M \log p(x_m|\theta)$ となる。このとき $m = (1, \dots, M)$ はテストセットのデータ点とし、 θ はモデルの推定パラメータとする。

4.2 人工データ実験

人工データ実験では NMF, VBNMF (NMF with sparse constraint) [4], PMF [15], そして NMMF の性能を比較する。また、スケールパラメータ 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、グリッドサーチを行なっている。人工データは、図 2 の確率的生成モデルによって生成した。X, Y, A, B のサイズは $I = J = N = M = 100$ とした。モデル性能とデータスパースネスの関係を検証するため、データのスパースネスを 0%, 0.9%, 9%, 99% の 4 段階に設定して実験を行った。

4.3 結果

人工データの実験結果を表 1 に示す。またスケールパラメータのグリッドサーチ結果を図 3 に示す。NMMF は、データのスパースネスが 9% 以上の場合において、他の手法よりも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。グリッドサーチの結果 (図 3) を確認すると、適切なスケールパラメータの選択による NMMF の性能向上が確認できる。ただし、NMMF はスケールパラメータのオーダーが一致していれば対数尤度は安定して上昇するため、スケールパラメータの選択に対してロバストであると言える。総じて、補助行列を用いることによって、NMMF は他手法に比べ、非ゼロ要素の推定精度が向上していると言える。

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。

4.4.1 Last.fm データ

Last.fm データセットは *Hetrec 2011*^{*1} 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。 $\mathbf{X} \in \mathbb{R}_+^{1,892 \times 17,632}$ 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。 $\mathbf{Y} \in \mathbb{R}_+^{11,946 \times 17,632}$ 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。 $\mathbf{Z} \in \mathbb{R}_+^{1,892 \times 1,892}$ を補助行列とする。NMF と NMMF の性能比較を、 \mathbf{X} の分解によって評価する。このデータはスパース性が高く、 $x_{i,j}$ の 0.25% (3,687 要素) のみが非零である。基底数 K は試行実験により $K = 20$ とした。

4.4.2 Together データ

ソーシャルキュレーションサービスである Together のデータ (収集期間は 2009 年 9 月から 2012 年 7 月まで) を利用する。ストーリーにおけるユーザの発言 (メッセージ) 頻度を *bags of users* として表現した観測行列を、 $\mathbf{X} \in \mathbb{R}_+^{1,823,184 \times 235,086}$ とする。次元はユーザに対応し、データ点はストーリーに対応する。 \mathbf{X} の要素 $x_{i,j}$ は非負値であり、全要素のうち非零要素は 0.0018% のスパースな行列となる。行列 \mathbf{X} は約 180 万 \times 20 万の大きさとなるが、NMMF の更新式はデータのスパース性を利用して、効率的に計算を行える。ストーリーに対する補助データとして、ストーリーでの単語頻度を *bags of words* として表現した、補助行列を $\mathbf{Y} \in \mathbb{R}_+^{165,046 \times 235,086}$ とする。次元は語彙種類に対応し、データ点はストーリーに対応する。 \mathbf{Y} は、非零要素が 1.23% のスパースな行列となる。ユーザ毎の補助データとして、Twitter でのフォロワー数とリスト数 \tilde{z} を、 $z = \log(\tilde{z} + 1)$ と変換した数値を、補助行列 $\mathbf{Z} \in \mathbb{R}_+^{1,823,184 \times 2}$ とする。基底数 K は試行実験により $K = 200$ とした。

4.5 結果

実データの定量評価の結果を表 2 に示す。2 つの実データセットにおいて NMMF は NMF に対して優位な数値を示している。表 2 から、NMMF の実データに対する優位性が確認できた。

5. 考察

本稿では、高度に疎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、Non-negative Multiple Matrix Factorization (NMMF) を提案した。NMMF で扱う問題を定式化したのち、一般化 KL 距離を用いて近似誤差を定め、補助関数法によるパラメータ推定法を提示した。同時に、NMMF のブロック未定義領域有り 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、NMMF と他の既存手法の性能を比較し、NMMF の優位性を示した。

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

Sparse	NMF	VBNMF	PMF	NMMF
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

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もつデータセットに対して, クロスバリデーションを行った. 平均値と標準偏差を表示した.

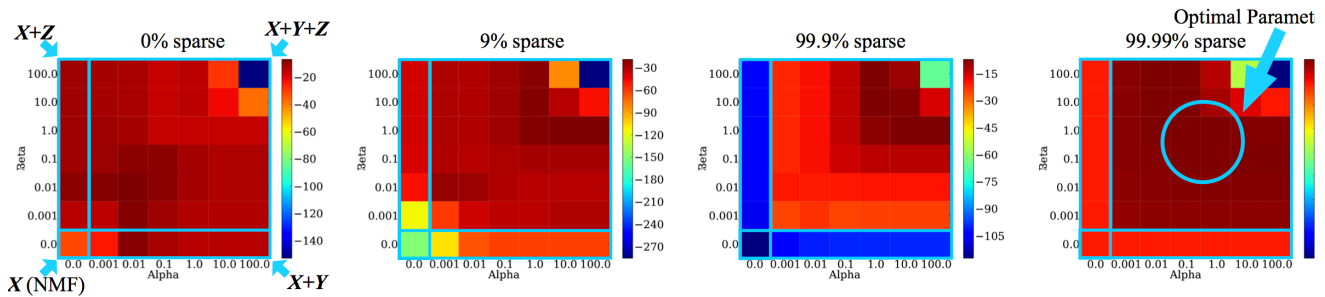


図 3 テスト対数尤度とスケールパラメータ α (水平軸) and β (鉛直軸) の関係

Data set	NMF	VBNMF	PMF	NMMF
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

表 2 実データ: テスト平均対数尤度によるモデルの性能比較.

今後は, より多数の行列を用いた場合の複合行列分解や, 非負制約を課さない複合行列分解の検討が必要である. 非負制約においては, 一般的な β 距離での定式化や, データごとに異なる β 距離を用いた定式化などを検討する必要がある. この際には, パラメータ推定が困難になることから, テクニカルな問題を解決していく必要があると考えられる.

参考文献

- [1] M. Aharon, M. Elad, and A.M. Bruckstein. On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them. *Linear algebra and its applications*, 416(1):48–67, 2006.
- [2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
- [3] B. Cao, D. Shen, J.T. Sun, X. Wang, Q. Yang, and Z. Chen. Detect and track latent factors with online nonnegative matrix factorization. In *Proc. IJCAI*, 2007.
- [4] A.T. Cemgil. Bayesian inference for nonnegative matrix factorisation models. *Computational Intelligence and Neuroscience*, 2009.
- [5] A. Cichocki, A. H. Phan R. Zdunek, and S. Amari. *Non-negative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis*, John Wiley. Wiley, 2009.
- [6] C. Ding, T. Li, and W. Peng. NMF and PLSI: Equivalence and a hybrid algorithm. In *Proc. SIGIR*, 2006.
- [7] T. Hofmann. Probabilistic latent semantic indexing. In *Proc. SIGIR*, 1999.
- [8] P.O. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning*

- Research*, 5:1457–1469, 2004.
- [9] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *IEEE*, 2009.
- [10] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, October 1999.
- [11] J. Lin, R. Snow, and W. Morgan. Smoothing techniques for adaptive online language models: topic tracking in tweet streams. In *Proc. SIGKDD*, 2011.
- [12] C. Liu, H. Yang, J. Fan, L. He, and Y. Wang. Distributed nonnegative matrix factorization for web-scale dyadic data analysis on MapReduce. In *Proc. WWW*, 2010.
- [13] H. Ma, H. Yang, M. R. Lyu, and I. King. Sorec: Social recommendation using probabilistic matrix factorization. In *Proc. CIKM*, 2008.
- [14] Q. Mei, D. Cai, D. Zhang, and C. Zhai. Topic modeling with network regularization. In *Proc. WWW*, 2008.
- [15] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In *Proc. NIPS*, 2008.
- [16] P. Smaragdis and J.C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *Proc. WASPA*, 2003.
- [17] C. Wang and D. M. Blei. Collaborative topic modeling for recommending scientific articles. In *Proc. SIGKDD*, 2011.
- [18] F. Wang, C. Tan, A.C. König, and P. Li. Efficient document clustering via online nonnegative matrix factorizations. In *Proc. SIAM*, 2011.
- [19] W. Xu, X. Liu, and Y. Gong. Document clustering based on non-negative matrix factorization. In *Proc. SIGIR*, 2003.