

特徴表現のスパース性を考慮したNMF

木村 圭吾¹ 吉田 哲也^{2,a)}

受付日 2011年8月16日, 再受付日 2011年9月26日,
採録日 2011年10月11日

概要: 本稿では, 特徴表現のスパース制約を考慮した NMF (Non-negative Matrix Factorization) を提案する. 近年, 要素が非負である実行列を, 同じく要素が非負である実行列の積として表現する非負値行列分解 (NMF) が注目を集めている. 従来研究では NMF における非負性制約が非零の要素が少ないスパースな特徴表現の学習に寄与すると考えられ, またスパース制約を導入した手法も提案されているが, これまで特徴表現のスパース性は明示的には考慮されてこなかった. 本稿では NMF における特徴表現に着目し, 特徴表現のスパース性を独立性と相関から定式化し, 定式化したスパース性を正則化項として活用する手法を提案する. 提案法を文書クラスタリングに適用し, 従来法との比較を通じて提案法の有効性を示す.

キーワード: スパース表現, 非負値行列分解, 正則化

Non-negative Matrix Factorization with Sparse Features

KEIGO KIMURA¹ TETSUYA YOSHIDA^{2,a)}

Received: August 16, 2011, Revised: September 26, 2011,
Accepted: October 11, 2011

Abstract: We propose an approach for Non-negative Matrix Factorization (NMF) with sparseness constraints on features. It has been believed that the non-negativity constraint in NMF contributes to making the learned features sparse. In addition, several approaches incorporated additional sparseness constraints, by hoping that the constraints make the features more sparse. However, previous approaches have mostly focused on coefficients, and have not considered the sparsity of features explicitly. Our approach explicitly incorporates the sparsity of features, in terms of independence of features and correlation of features. The proposed notion of sparsity is formalized as regularization terms in the framework of NMF, and learning algorithms with multiplicative update rules are proposed. The proposed approach is evaluated in terms of document clustering over well-known benchmark datasets. Several experiments have been conducted on the datasets, and comparison with other state-of-the-art NMF algorithms is reported. The results are encouraging and show that the proposed approach improves the clustering performance, while sustaining relatively good quality of data approximation.

Keywords: sparse coding, Non-negative Matrix Factorization, regularization

1. はじめに

本文中の単語頻度や生体の視覚野におけるニューロンからの活動電位など, 観測データは非負のデータ行列として

表現されることが多い. このため, 非負の信号の重ね合わせに基づく新しい情報処理のモデルとして Non-negative Matrix Factorization (NMF) の研究が活発に行われている [1], [4], [5], [6], [9], [13]. 主成分分析などの固有値解析に基づく手法と異なり, NMF ではデータの特徴ベクトル (特徴表現) の線形結合として近似的に分解する際に特徴ベクトルが相互に直交するとは限らず, 視覚野などの研究 [5], [9] や画像処理などへの適用例が報告されている [3], [4]. また, 文書クラスタリングに対する適用例も報告されてい

¹ 北海道大学工学部
Faculty of Engineering, Hokkaido University, Sapporo,
Hokkaido 060-8628, Japan

² 北海道大学大学院情報科学研究科
Graduate School of Information Science and Technology,
Hokkaido University, Sapporo, Hokkaido 060-0814, Japan

a) yoshida@meme.hokudai.ac.jp

る [13].

従来の研究においては、要素が非負である実行列を、同じく要素が非負である実行列の積として近似的に表現する際、NMFにおける非負性制約が非零の要素が少ないスパースな特徴表現の学習に寄与すると考えられてきた。また、行列に対するスパース制約を導入し、よりスパースな特徴表現の学習を試みる手法も提案されている [4], [7]。しかし、NMFで学習する特徴表現のスパース性はこれまで明示的には考慮されてこなかった。

本稿では、非負の表現を得るためにNMFを用いる際、NMFで学習する特徴表現に着目し、特徴表現のスパース性を独立性と相関から定式化し、定式化したスパース性を正則化項として活用する手法を提案する。提案する正則化項を追加した目的関数を定義し、この目的関数を最小化するアルゴリズムを提案する。さらに、提案アルゴリズムの収束性を示す。

提案法を文書クラスタリングに適用して評価し、従来法との比較を通じて提案法の有効性を報告する。提案法により特徴表現のスパース性を考慮した行列分解を行うことにより、行列分解における近似精度を保ちながらクラスタリング性能を向上させることができることを報告する。

2章では関連研究を紹介し、3章で提案法について説明し、4章で評価実験の結果を報告し、提案法の有効性を議論する。5章でまとめと今後の展望を述べる。

1.1 準備

本稿では、行列は太字の大文字、ベクトルは太字のイタリック小文字で表記し、 \mathbf{X}_{ij} で行列 \mathbf{X} の第 ij 要素を表す。tr は行列のトレースを表し、 \mathbf{X} の転置を \mathbf{X}^T で表す。 \mathbf{I} は単位行列、 $\mathbf{1}$ は要素がすべて1のベクトルを表し、正則行列 \mathbf{A} の逆行列を \mathbf{A}^{-1} で表す。与えられたデータが表現される空間をデータ空間、写像先の空間を特徴空間と呼ぶ。データ空間を構成するベクトルを属性、特徴空間を構成するベクトルを特徴と呼び、ベクトルの各要素は非負とする。

2. 関連研究

2.1 Non-negative Matrix Factorization (NMF)

Non-negative Matrix Factorization (NMF) [6] とは、与えられた p 個の属性で表現される非負のデータ行列 $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}_+^{p \times n}$ (n はデータ数) を*1、特徴数 q のもとで非負の行列 $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_q] \in \mathbb{R}_+^{p \times q}$ 、 $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n] \in \mathbb{R}_+^{q \times n}$ の積により

$$\mathbf{X} \simeq \mathbf{UV} \quad \text{ただし、} \mathbf{U}, \mathbf{V} \text{ は非負行列} \quad (1)$$

と近似して分解する行列 \mathbf{U}, \mathbf{V} を求める手法である。行列 \mathbf{U} は特徴行列、 \mathbf{V} は係数行列と呼ばれる。NMFにより

*1 本稿では、NMFにおける記法に従い、 $\mathbf{X} \in \mathbb{R}_+^{p \times n}$ の形式をデータ行列と呼ぶ。

各データ \mathbf{x}_i は $\mathbf{u}_1, \dots, \mathbf{u}_q$ の線形結合として近似的に $\mathbf{U}\mathbf{v}_i$ と表現されることになる。NMFはデータ行列 \mathbf{X} と行列積 \mathbf{UV} との差異の最小化を通じて行列分解を行い、式 (1) での近似の良さを表現する以下の目的関数の最小化を行う。

$$J_0 = \|\mathbf{X} - \mathbf{UV}\|^2 \quad (2)$$

ここで $\|\cdot\|$ は行列のノルムであり、フロベニウスノルムとKL情報量を用いる方法が提案されているが [6]、本稿ではフロベニウスノルムを用いる場合を扱う。

NMFを次元削減法と見なす場合には、行列 \mathbf{U} の列ベクトルを特徴と見なし、係数行列 \mathbf{V} を \mathbf{U} のもとでの低次元データ表現として扱う。このため、 p 次元列ベクトルであるデータ \mathbf{x}_i はNMFを通じて q 次元列ベクトルである \mathbf{v}_i として表現される。データ処理の際には、NMFにより求めた低次元表現である \mathbf{V} に対して既存の手法を適用することが多い。

2.2 スパース制約を考慮したNMFアルゴリズム

NMFにより低次元データ表現である \mathbf{V} を求める際には非負性制約が非零の要素が少ないスパースな特徴表現の学習に寄与すると考えられてきたが、分解時に行列に対してスパース制約を導入し、よりスパースな特徴表現の学習を試みる手法も提案されている [4], [7]。

文献 [7] は係数行列 \mathbf{V} に対してスパース制約を導入する手法を提案した。スパース制約を行列 \mathbf{V} の列ベクトルごとのコストとして表現し、以下の目的関数の最小化を行う。

$$J_{SNMF} = \|\mathbf{X} - \mathbf{UV}\|^2 + \nu \sum_{i=1}^n \|\mathbf{v}_i\|_1 \quad (3)$$

ここで $\|\cdot\|_1$ は L_1 ノルムであり [10]、 ν はパラメータである。式 (3) の最小化に基づき、文献 [6] での行列 \mathbf{U}, \mathbf{V} の更新式 [6] に類似した更新式が提案された。しかし、式 (1) でのデータ行列の近似の際、行列 \mathbf{U} と \mathbf{V} のスケールに関する不定性*2が存在することが知られている [13]。式 (3) における正則化項 (第2項) は係数行列 \mathbf{V} の各要素の絶対値を一様に小さくすることに寄与するが、データ行列の近似においてこの影響を補填するために行列 \mathbf{U} の要素が相対的に大きくなるにとどまる [7]。

文献 [4] はNMFにおいてスパースな行列を得るために非線形射影に基づく手法を提案した。まず文献 [6] での行列 \mathbf{U}, \mathbf{V} の更新式 [6] を適用し、次に指定されたスパース性に応じた非線形射影を用いて行列の列ベクトルごとに更新を行い、負値となった要素は非負性を保証するためにさらに零に設定する。この手法では行列 \mathbf{U} と \mathbf{V} は同様に処理できるが、事前に指定したスパース性に依りてベクトルごとに非線形射影を通じたスパース化を行う。ベクトルの

*2 $\mathbf{X} \simeq \mathbf{UV}$ と近似された場合、任意の正則行列 \mathbf{S} を用いて同様に $\mathbf{X} \simeq \mathbf{US}^{-1}\mathbf{SV}$ と近似できる。

次元数を r とした場合、スパース性の程度は下記で定義される。

$$sp = \frac{\sqrt{r} - (L_2/L_1)}{\sqrt{r} - 1} \quad (4)$$

ここで L_1 は式 (3) と同様にベクトルの L_1 ノルムであり、 L_2 は L_2 ノルムを表す。事前に指定した式 (4) の値に応じた処理を行うため、与えられたデータの性質を反映したスパースさを得られるわけではない。

3. 特徴表現のスパース性を考慮した NMF

本稿では、特徴表現に対するスパース制約を独立性と相関から考慮する手法を提案する。それぞれの詳細を 3.1 節と 3.2 節で述べる。

3.1 特徴の独立性

3.1.1 独立性に基づく正則化

NMF によりデータ行列 \mathbf{X} を近似する際、各データ \mathbf{x}_i は $\mathbf{u}_1, \dots, \mathbf{u}_q$ の線形結合として近似的に表現されるため、各データは行列 \mathbf{U} の列ベクトルで張られる特徴空間において表現されることになる。しかし、NMF の非負性により一般に行列 \mathbf{U} の列ベクトルはお互いに直交するとは限らず、特徴数 q が多い場合には線形従属となることが多い。

行列 \mathbf{U} の列ベクトルは NMF で得られる特徴空間における座標系の役割を果たすため、線形独立となっていることが望ましいと考えられる。逆に、特徴数を多くしても線形独立でなく冗長な特徴が多く存在する場合には、行列 \mathbf{U} の列ベクトルで構成される特徴空間が狭くなり与えられたデータ行列を十分に近似できない恐れがある。

本稿では行列 \mathbf{U} の各列ベクトル \mathbf{u}_i に対する正規化条件 $\mathbf{u}_i^T \mathbf{u}_i = 1$ のもとで [13] 特徴どうしの独立性を $\mathbf{1}_q^T \mathbf{U}^T \mathbf{U} \mathbf{1}_q$ と定式化し*3、下記の目的関数の最小化を提案する。

$$J_1 = \|\mathbf{X} - \mathbf{UV}\|^2 + \lambda_1 \mathbf{1}_q^T \mathbf{U}^T \mathbf{U} \mathbf{1}_q \quad (5)$$

ここで $\lambda_1 \geq 0$ はパラメータである。上記の正規化条件のもとでは $\mathbf{U}^T \mathbf{U}$ の対角要素はすべて定数であるため、非負性により NMF で学習する特徴どうしが線形独立になるほど $\mathbf{U}^T \mathbf{U}$ の非対角要素が零に近づくため式 (5) の第 2 項は小さくなる。

3.1.2 更新式

式 (5) の最小化により \mathbf{U} , \mathbf{V} を求めるアルゴリズムを提案する。 \mathbf{U} , \mathbf{V} の各要素の非負性制約に対応するラグランジュ未定係数 Ψ , Φ を導入して式 (6) を定義する。

$$\mathcal{L} = \|\mathbf{X} - \mathbf{UV}\|^2 + \lambda_1 \mathbf{1}_q^T \mathbf{U}^T \mathbf{U} \mathbf{1}_q + \text{tr}(\Psi \mathbf{U}^T) + \text{tr}(\Phi \mathbf{V}^T) \quad (6)$$

式 (6) を \mathbf{U} , \mathbf{V} に対して微分し、下記を得る。

*3 1.1 節で述べたように $\mathbf{1}_q$ は要素がすべて 1 の q 次元ベクトルを表す。

```

Require:  $\mathbf{X} \in \mathbb{R}_+^{p \times n}$  //data matrix
Require:  $q \in \mathbb{N}$  //number of features
Require:  $\lambda_1$  //hyper-parameter
1: initialize  $\mathbf{U}$ ,  $\mathbf{V}$ 
2: while not yet terminated do
3:    $\mathbf{U}_{ij} \leftarrow \mathbf{U}_{ij} \frac{(\mathbf{X}\mathbf{V}^T)_{ij}}{(\mathbf{U}\mathbf{V}\mathbf{V}^T + \lambda_1 \mathbf{U}\mathbf{1}_q \mathbf{1}_q^T)_{ij}}$ 
4:    $\mathbf{V}_{ij} \leftarrow \mathbf{V}_{ij} \sqrt{\sum_i \mathbf{U}_{ij}^2}$ 
5:    $\mathbf{U}_{ij} \leftarrow \frac{\mathbf{U}_{ij}}{\sqrt{\sum_i \mathbf{U}_{ij}^2}}$ 
6:    $\mathbf{V}_{ij} \leftarrow \mathbf{V}_{ij} \frac{(\mathbf{U}^T \mathbf{X})_{ij}}{(\mathbf{U}^T \mathbf{U}\mathbf{V})_{ij}}$ 
7: end while
8: return  $\mathbf{U}$ ,  $\mathbf{V}$ 
    
```

図 1 アルゴリズム scNMF

Fig. 1 Algorithm scNMF.

$$\frac{\partial \mathcal{L}}{\partial \mathbf{U}} = -2\mathbf{X}\mathbf{V}^T + 2\mathbf{U}\mathbf{V}\mathbf{V}^T + 2\lambda_1 \mathbf{U}\mathbf{1}_q \mathbf{1}_q^T + \Psi \quad (7)$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{V}} = -2\mathbf{X}^T \mathbf{U} + 2\mathbf{V}^T \mathbf{U}^T \mathbf{U} + \Phi^T \quad (8)$$

KKT (Karush-Kuhn-Tucker) 条件 [14] より各要素ごとに以下が成立する。

$$(-\mathbf{X}\mathbf{V}^T + \mathbf{U}\mathbf{V}\mathbf{V}^T + \lambda_1 \mathbf{U}\mathbf{1}_q \mathbf{1}_q^T)_{ij} (\mathbf{U})_{ij} = 0 \quad (9)$$

$$(-\mathbf{X}^T \mathbf{U} + \mathbf{V}^T \mathbf{U}^T \mathbf{U})_{ij} (\mathbf{V})_{ij}^T = 0 \quad (10)$$

式 (9), (10) に基づき、以下の更新式が得られる。

$$(\mathbf{U})_{ij} \leftarrow (\mathbf{U})_{ij} \frac{(\mathbf{X}\mathbf{V}^T)_{ij}}{(\mathbf{U}\mathbf{V}\mathbf{V}^T + \lambda_1 \mathbf{U}\mathbf{1}_q \mathbf{1}_q^T)_{ij}} \quad (11)$$

$$(\mathbf{V})_{ij} \leftarrow (\mathbf{V})_{ij} \frac{(\mathbf{U}^T \mathbf{X})_{ij}}{(\mathbf{U}^T \mathbf{U}\mathbf{V})_{ij}} \quad (12)$$

3.1.3 アルゴリズム scNMF

式 (11), (12) の更新式に基づく提案アルゴリズム scNMF (sparsity constrained NMF) を図 1 に示す。1 行目で行列を初期化し、3, 6 行目で \mathbf{U} , \mathbf{V} の交互最適化を行う。2 行目の終了条件として繰返し数や式 (2) の値が十分小さい場合とすることが多い。2.2 節で述べたスケールに関する不定性に対処するため、4, 5 行目で文献 [13] での正規化を行う。5 行目は提案法における式 (5) での正規化条件を実現していることに注意されたい。

scNMF について以下の性質が成り立つ。

定理 1. 式 (5) はアルゴリズム scNMF の更新式により単調非増加であり、非負であるためアルゴリズムは収束する。

収束性の証明は付録に示す。

3.2 特徴の相関

3.2.1 相関に基づく正則化

2.1 節で述べたように、NMF は指定された p 個の属性で表現されるデータ行列を分解する。NMF で学習する特徴 $\mathbf{u}_1, \dots, \mathbf{u}_q$ を用いてデータを特徴空間において近似的に表

現する際には、データ空間における構造を保存することが望ましいと考えられる。本稿では、データ空間における構造をデータ空間における属性間での類似性から考慮する。

データ空間における属性間での類似性を保存するため、対関係を表現するグラフ構造を構築し、グラフに基づく正則化法を活用する [1], [8]。ただし、与えられたデータ空間における関係を考慮するために、データ対 [1] や特徴対 [8] ではなく、本稿ではデータ空間における属性対の関係を考慮する。

データ空間における属性対の類似度が与えられた場合、類似度を辺の重みとする無向グラフとして属性対の関係を表現する。 $\mathbf{W} \in \mathbb{R}_+^{p \times p}$ を類似度 (本稿ではコサイン類似度を用いた) に基づく m -近傍グラフ^{*4}の重み行列として、以下の目的関数の最小化を考える：

$$J_2 = \|\mathbf{X} - \mathbf{UV}\|^2 + \lambda_1 \mathbf{1}_q^T \mathbf{U}^T \mathbf{U} \mathbf{1}_q + \lambda_2 \text{tr}(\mathbf{U}^T \mathbf{L} \mathbf{U}) \quad (13)$$

$\mathbf{L} = \mathbf{D} - \mathbf{W}$ はグラフの次数行列 \mathbf{D} ^{*5} に基づくグラフラプラシアンである [12]。第 3 項がデータ空間における属性に基づく正則化項であり、 $\lambda_2 \geq 0$ はパラメータである。トレースの交代性から $\text{tr}(\mathbf{U}^T \mathbf{L} \mathbf{U}) = \text{tr}(\mathbf{U} \mathbf{U}^T \mathbf{L})$ であり、 $\mathbf{U} \mathbf{U}^T$ は NMF で学習する特徴行列 \mathbf{U} の相関に対応するため、提案法はグラフ構造に基づいて特徴の相関を活用することになる。

3.2.2 更新式

3.1.2 項と同様に、ラグランジュ未定係数に基づく停留条件および KKT 条件を用いて式 (13) を最小化する以下の更新式を導出する。式 (13) で行列 \mathbf{V} に依存するのは第 1 項のみであるため、 \mathbf{V} に対しては式 (12) と同じ更新式となる。他方、行列 \mathbf{U} に対しては以下の更新式となる：

$$(\mathbf{U})_{ij} \leftarrow (\mathbf{U})_{ij} \frac{(\mathbf{X} \mathbf{V}^T + \lambda_2 \mathbf{W} \mathbf{U})_{ij}}{(\mathbf{U} \mathbf{V} \mathbf{V}^T + \lambda_1 \mathbf{U} \mathbf{1}_q \mathbf{1}_q^T + \lambda_2 \mathbf{D} \mathbf{U})_{ij}} \quad (14)$$

3.2.3 アルゴリズム scGNMF

式 (13) を最小化するアルゴリズム scGNMF では、図 1 の 3 行目の更新式の代わりに式 (14) の更新式を用いる。定理 2 と同様に、提案アルゴリズムについて以下の性質が成り立つ。

定理 2. 式 (13) は上記の更新式により単調非増加であり、非負であるためアルゴリズムは収束する。

収束性の証明は付録に示す。

4. 評価

4.1 実験設定

4.1.1 対象データ

提案法を単語の頻度に基づくベクトル空間モデルで表

^{*4} 各頂点ごとに類似度の降順に m 個の他の頂点と接続して構築される。

^{*5} 行列 \mathbf{W} に対し、 $d_i = \sum_j (\mathbf{W})_{ij}$ とするベクトル \mathbf{d} に基づく対角行列 $\mathbf{D} = \text{diag}(\mathbf{d})$ を次数行列と呼ぶ。

現された文書データに適用して文書クラスタリングを行い、従来法との比較を行った。使用したデータセットは 20 ニュースグループ (以下、20NG と表記)^{*6}, 2) SRAA^{*7}, 3) TREC^{*8} である。文章における単語頻度は非負であり、テキストモデル [11] などのように文書をいくつかの話題の混合と見なす場合、それぞれの話題での単語の生起頻度や話題の含まれ方は非負であるため、非負制約に基づく NMF を適用して解析することは自然なアプローチであると考えられる。

20NG に対して 5 クラスタ、10 クラスタ、15 クラスタからなる 3 つの母集団を設定し、各クラスタからそれぞれ 50 個ずつの文書を非復元抽出してデータセットを作成した。各母集団に含まれるニュースグループを表 1 に示す。各母集団に対して 10 個ずつのサンプルを作成した。各サンプルごとに porter stemmer^{*9} を用いて stemming を行い、MontyTagger^{*10} を用いて品詞に分解し、stop word を除去して相互情報量で上位 2,000 語の単語を選択した。

SRAA は UseNet における 4 つのカテゴリでの投稿記事を集めたものである。73,218 の記事から本文を抜き出し、短文の記事を除外するために各カテゴリごとにファイルサイズで上位 500 個の記事を抽出し、20NG と同様な処理を行い 10 個のサンプルを作成した。

TREC に対しては表 2 に示すデータセットを使用した。上記と同様な処理により各データセットに対して 10 個ずつのサンプルを作成したが、TREC ではすでに前処理済みであるため個別の前処理などは行わなかった。

4.1.2 評価尺度

上記のデータはデータ (ここでは文書) ごとに真のクラスタ (ラベル) が既知であるため、データのラベルに基づき正規化相互情報量 (NMI) を評価した。真のクラスタと割り当てられたクラスタに対応する確率変数を C, \hat{C} とすると、NMI は以下で定義される。

$$NMI = \frac{I(\hat{C}; C)}{(H(\hat{C}) + H(C))/2} \quad (\in [0, 1]) \quad (15)$$

$H(\cdot)$ はシャノン情報量であり、 $I(\cdot; \cdot)$ は相互情報量である。NMI が大きいほど真のクラスタでのデータ割当てに合致することを示すため、クラスタ割当ての精度に対応する。

NMF とは与えられたデータ行列 \mathbf{X} を \mathbf{UV} として近似して分解する手法であるため、分解における (正規化した) 近似精度を以下で測る。

$$\text{approx.} = \frac{\|\mathbf{X} - \mathbf{UV}\|}{\|\mathbf{X}\|} \quad (16)$$

式 (16) の値が小さいほど行列 \mathbf{U}, \mathbf{V} による近似が良い

^{*6} <http://people.csail.mit.edu/~jrennie/20Newsgroups/> (本稿では 20news-18828 を使用)

^{*7} <http://www.cs.umass.edu/~mccallum/data.html>

^{*8} <http://glaros.dtc.umn.edu/gkhome/cluto/cluto/download>

^{*9} <http://www.tartarus.org/~martin/PorterStemmer>

^{*10} <http://web.media.mit.edu/~hugo/montytagger>

表 1 20 Newsgroup に対するデータセット
Table 1 Datasets from 20 Newsgroup dataset.

データセット	含まれるグループ名
Multi5	comp.graphics, rec.motorcycles, rec.sport.baseball, sci.space, talk.politics.mideast
Multi10	alt.atheism, comp.sys.mac.hardware, misc.forsale, rec.autos, rec.sport.hockey, sci.crypt, sci.med, sci.electronics, sci.space, talk.politics.guns
Multi15	alt.atheism, comp.graphics, comp.sys.mac.hardware, misc.forsale, rec.autos, rec.motorcycles, rec.sport.baseball, rec.sport.hockey, sci.crypt, sci.electronics, sci.med, sci.space, talk.politics.guns, talk.politics.mideast, talk.politics.misc

表 2 TREC データセット

Table 2 TREC datasets (original representation).

dataset	# attr.	#class	#data
sports	126,372	7	8,580
reviews	126,372	5	4,069
k1b	21,839	6	2,340
ohscal	11,465	10	11,162

ことに対応する.

また, 文献 [4] で提案された式 (4) の指標が NMF におけるスパース性として他の文献でも広く用いられているため, 特徴行列のスパース性を式 (4) の指標で評価した.

4.1.3 比較手法

提案法を, 代表的な NMF アルゴリズムである下記の手法と比較した.

- (a) スパース制約を考慮しない手法: 1) NMF [6], 2) WNMf [13], 3) GNMF [1]
- (b) スパース制約を考慮した手法: 4) SNMF [7], 5) NMFsc [4]
- (c) 特徴行列 \mathbf{U} を考慮した手法: 6) ONMF [2]

Weighted NMF (WNMF) とは, Ncut 法 [12] における重み付けを活用してデータ表現を変換し, 変換後の表現に対して NMF を適用する手法である [13]. データ行列 \mathbf{X} に対するグラム行列 $\mathbf{X}^T\mathbf{X}$ を重み行列 \mathbf{W} とするグラフを考え, \mathbf{W} に対する次数行列 \mathbf{D} の逆行列を Γ としてデータ行列 \mathbf{X} を $\mathbf{X}\Gamma^{1/2}$ に変換し, $\mathbf{X}\Gamma^{1/2}$ に対して通常の NMF [6] を適用する.

Graph regularized NMF (GNMF) とは, 3.2 節と同様に m -近傍グラフを用いる手法であるが, GNMF ではデータ対に関する m -近傍グラフを構築し, その隣接行列 \mathbf{A} に対するグラフラプリアン $\mathbf{L} = \mathbf{D} - \mathbf{A}$ を用いた以下の目的関数の最小化を行う [1].

$$J_3 = \|\mathbf{X} - \mathbf{UV}\|^2 + \lambda \text{tr}(\mathbf{V}\mathbf{L}\mathbf{V}^T) \quad (17)$$

ここで λ はパラメータである.

Orthogonal NMF (ONMF) [2] とは直交制約のもとで行列分解を行う手法である. 3.1 節で述べた特徴の独立性は行列 \mathbf{U} に対する直交制約 ($\mathbf{U}^T\mathbf{U} = \mathbf{I}$) を緩和し, 逆に $\mathbf{U}^T\mathbf{U}$ の非対角要素への制約を課したものと見なすことができるため, ONMF との比較を行った.

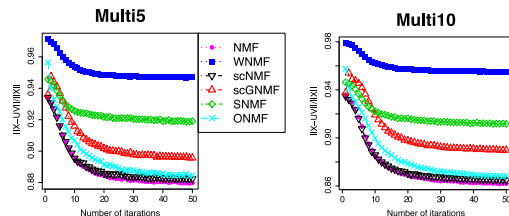


図 2 交互最適化での反復回数の影響

Fig. 2 Influence of iterations.

4.1.4 実験手順

各手法で構築した係数行列 \mathbf{V} に対し, 各データセットのクラスタ数 k を与えて skmeans 法を適用してクラスタリングを行った. NMF は局所最適化を行う手法であるため, 得られる結果 (行列 \mathbf{U}, \mathbf{V}) は初期値に依存する. このため, 1つのデータ行列に対して行列 \mathbf{U}, \mathbf{V} の初期値をランダムに 10 回変えて実験した. さらに, 各データセットで 10 個のサンプルに同様な処理を行い, その平均を求めた*11.

4.1.5 実験パラメータ

対関係 (データ対, 特徴対, 属性対) における類似度には文書処理で標準的なコサイン類似度を用いた. 類似度が上位 10 個の近傍を用いて m -近傍グラフを構築し ($m = 10$), GNMF に対する式 (17) での λ は文献 [1] で推奨された値 ($\lambda = 100$) とした. 文献 [7] に従い SNMF におけるパラメータ ν を 100 とし, NMFsc における式 (4) の sp の値は 0.2 とした*12.

交互最適化による更新における実験条件をそろえるため, すべての手法で図 1 の 2 行目での終了条件を反復回数とした. 特徴数 q をクラスタ数として反復回数の影響を調べた結果を図 2 に示す (横軸は反復回数, 縦軸は式 (16) の近似誤差, 凡例は 4.2.1 項参照). 図 2 より反復回数が 30 回ほどでほぼ収束しているため, 以下では反復回数を 30 回とした.

4.1.6 提案法におけるパラメータ値の感度評価

提案法におけるパラメータは式 (13) における λ_1, λ_2 である. この影響を調べるため, 特徴数 q を固定し, パラ

*11 サンプルごとに 10 回試行したため, 計 100 回の試行の平均を報告する.

*12 スパースな行列を得るために式 (4) の値を小さくした場合には 0 のみの列ベクトルが生成され停止してしまうことがあるため, 予備実験に基づき 0.2 とした.

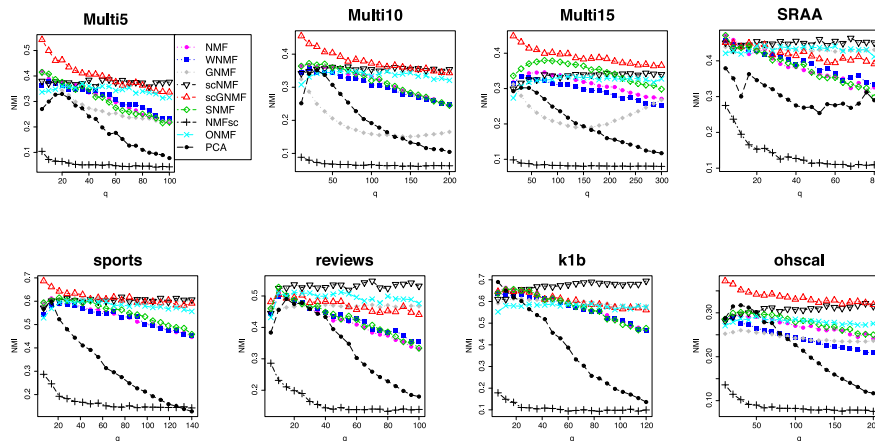


図 4 クラスタリング精度 (NMI)
Fig. 4 Clustering performance (NMI).

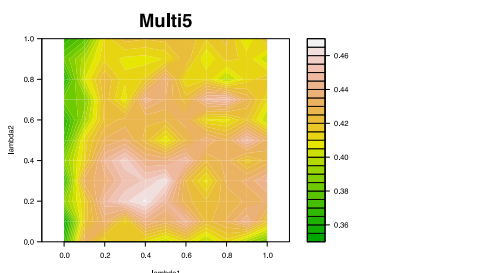


図 3 パラメータの影響 (NMI)
Fig. 3 Influence of parameters.

メータ値を変えて実験した結果を図 3 に示す (Multi5 で特徴数 $q = 10 \times k$ の場合を示す)。図 3 で横軸は λ_1 、縦軸は λ_2 に対応し、右横の凡例にクラスタリング精度 (NMI) を示す。図 3 で特徴の独立性を考慮しない場合 ($\lambda_1 = 0$) よりも考慮した場合の方が精度 (NMI) が高く、提案法は効果的であることが分かる。同様に特徴の相関を考慮した場合 ($\lambda_2 > 0$) の方が精度は高くなったが、 λ_1 との組合せに応じて多少変動した。他の手法と同様、提案法もパラメータ値に影響を受けるが、以下の実験では λ_1 と λ_2 はそれぞれ 0.4 とした。

4.2 結果

2.1 節で述べたように NMF は指定された特徴数 q のもとの行列分解を行う。以下では特徴数 q をクラスタ数 k の $1, \dots, 20$ 倍まで変えて実験した。

4.2.1 クラスタリング精度

クラスタリング精度を図 4 に示す (横軸は特徴数 q 、縦軸は NMI)。凡例で破線三角は提案法 (黒は scNMF、赤は scGNMF) である。点線は 4.1.3 項の (a) にあげた手法であり、丸は NMF、四角は WNMf、菱形は GNMF である。(b) にあげた SNMF は菱形、NMFsc は十字で示す。(c) にあげた ONMF は×印で示す。また、主成分分析で得られる主成分得点をデータ表現とし、他手法と同様に skmeans

を適用した結果 (PCA) を黒丸で示す。

図 4 の結果より提案法 (scNMF と scGNMF) は精度の観点から他手法を上回った。特徴数 q が小さい場合には概して scGNMF の精度が高かったが、 q の増加にともない精度が悪化した。他方、scNMF の精度は q に対してあまり変化がなく、SRAA, reviews, k1b データセットでは scGNMF をも上回った。また、NMFsc 以外はおおむね PCA を上回った。

4.1.3 項の (b) にあげたスパース制約を考慮した従来法のうち、SNMF の性能は NMF とほぼ同程度であった。式 (3) の第 2 項は行列 V の要素を小さくするが、全要素を一様に小さくするためクラスタリングには影響しないためと考えられる。他方、NMFsc の精度は非常に低かった。これは、NMFsc では事前に設定した式 (4) の値に応じてデータの性質を考慮せずに非線形写像を通じたスパース化が行われるためと考えられる。

4.2.2 行列分解による近似精度

式 (16) の行列分解による近似精度を図 5 に示す (縦軸が近似精度で、小さいほど近似が良い)*13。図 5 の結果より、4.1.3 項の (b) にあげたスパース制約を考慮した手法以外では特徴数 q が増えるほど近似精度が向上した。特徴数 q が多いほど行列分解での自由変数が増えるために自然な結果である。従来の NMF には及ばなかったが、近似精度の観点からは提案法は (a) の WNMf (および GNMF)、(b) の SNMF, NMFsc を上回った。scGNMF は ONMF には及ばなかったが、scNMF は ONMF とほぼ同程度の近似精度を示した。

他方、(b) のスパース制約を考慮した従来法の近似精度は非常に悪かった。SNMF における L_1 正則化項や NMFsc における非線形射影など、これらの手法では与えられたデータの性質よりも分解時に行列の要素値を小さくするこ

*13 GNMF では式 (17) の第 2 項が非常に大きいため式 (16) の値が 100 以上となり、非常に近似精度が悪いために図 5 では省略した。

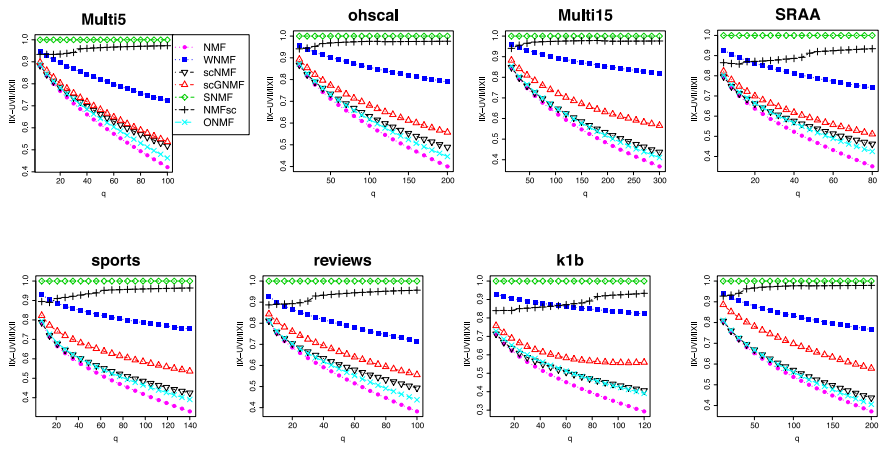


図 5 データ行列の近似精度

Fig. 5 Normalized quality of approximation in Eq. (16).

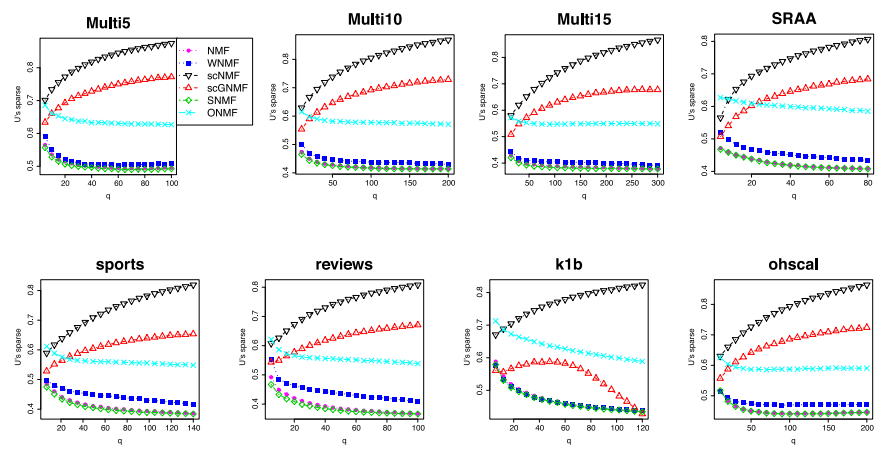


図 6 特徴行列のスパース性

Fig. 6 Sparsity of features.

とに主眼を置いているためと考えられる。

4.2.3 スパース性

本稿では特徴表現のスパース性に着目し、独立性と相関の観点からスパースな特徴行列を得る手法を提案した。提案法の効果を検証するため、特徴行列に対して式 (4) の指標を評価した結果を図 6 に示す。図 6 より、提案法では他手法よりもスパースな特徴が得られており、また特徴数 q が増えるほどよりスパースとなっていることが分かる。

4.2.4 特徴の独立性

4.1.3 項で述べたように、提案法における特徴の独立性は ONMF [2] における制約を汎化したものと見なせる。提案法と ONMF を特徴の独立性の観点から比較するため、それぞれの手法 (および NMF [6]) で学習した特徴における独立性 (式 (5) における $\mathbf{1}_q^T \mathbf{U}^T \mathbf{U} \mathbf{1}_q$) を評価した。結果を図 7 に示す (縦軸が独立性: 小さいほど独立性が高い)。

図 7 より、行列 \mathbf{U} に対する制約を用いることにより従来の NMF [6] と比較して提案法と ONMF では独立性が高いことが分かる。また、提案法は ONMF での制約を汎化したものに対応するため、scNMf の方が ONMF より独立性が高かった。他方、scGNMf では式 (13) の第 3 項の影響

により、特徴数 q が少ない場合には逆に ONMF よりも独立性が低くなったが、 q が増加するにつれ両者はほぼ同等な結果を示した。

4.3 議論

本稿では NMF におけるデータ行列の近似分解に対し、行列分解における近似精度に加えて特徴表現のスパース性を考慮することを提案した。提案法では特徴表現のスパース性を独立性と相関から定式化し、定式化したスパース性を NMF における目的関数に対する正則化項として活用する。4.2 節の実験結果より、提案法 (scNMf と scGNMf) は行列分解における近似精度を保ちながら従来法を上回るクラスタリング精度を示した。この結果から提案法は効果的であるといえる。

クラスタリングではラベル情報を活用できないため精度向上を保証する目的関数を設計して最適化を行うことは困難であるが、提案法を通じて独立性が高く共起しやすい属性を持つ特徴を構築してデータを近似表現したことが精度向上につながったと考えられる。

本稿で提案した特徴の独立性は ONMF [2] における (行

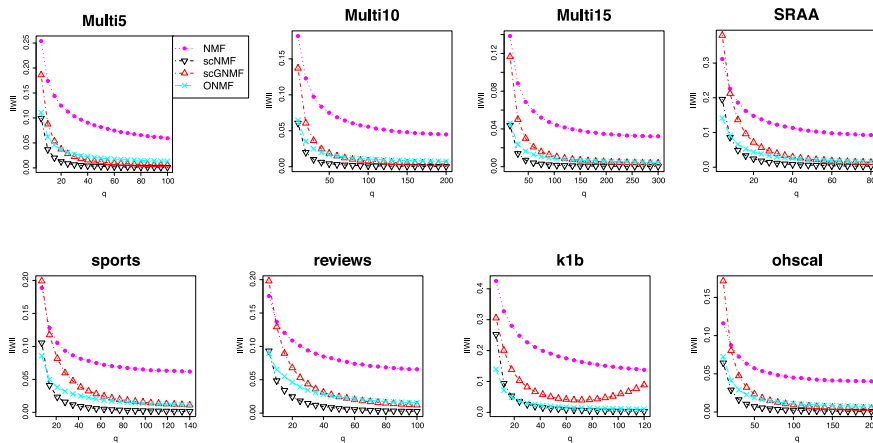


図 7 特徴の独立性

Fig. 7 Orthogonality of features.

列 \mathbf{U} に対する) 制約を汎化したものと見なすことができるが、提案法は特徴の独立性や近似精度の観点からは ONMF とほぼ同等な性能を示し、精度の観点からは上回る性能を示した。

5. おわりに

本稿では NMF で学習する特徴表現に着目し、特徴表現のスパース性を独立性と相関から定式化し、定式化したスパース性を正則化項として活用する手法を提案した。スパース性を考慮した従来手法のように行列分解において行列の要素値を小さくすることを目指すのではなく、NMF における行列分解に基づくデータ近似に着目して特徴表現に焦点を当てるアプローチを提案した。提案する正則化項を追加した目的関数を定義し、この目的関数を最小化するアルゴリズムを提案した。さらに、提案アルゴリズムの収束性を示した。

提案法を文書クラスタリングに適用して評価し、従来手法との比較を通じてその有効性を確認した。特に、提案法により特徴表現のスパース性を考慮することにより、行列分解における近似精度を保ちながらクラスタリング性能が向上することを確認した。今後は画像などの他の実データに対する評価を通じてさらなる改良を行う予定である。

参考文献

- [1] Cai, D., He, X., Wu, X. and Han, J.: Non-negative Matrix Factorization on Manifold, *Proc. ICDM'08*, pp.63-72 (2008).
- [2] Ding, C., Li, T., Peng, W. and Park, H.: Orthogonal Nonnegative Matrix Tri-factorizations for Clustering, *Proc. KDD '06*, pp.126-135 (2006).
- [3] Gong, D., Zhao, X. and Yang, Q.: Sparse Non-negative Pattern Learning for Image Representation, *Proc. ICIP*, pp.981-984 (2008).
- [4] Hoyer, P.O.: Non-negative Matrix Factorization with Sparseness Constraints, *Journal of Machine Learning Research*, Vol.5, pp.1457-1469 (2004).

- [5] Lee, D.D. and Seung, H.S.: Learning the parts of objects by non-negative matrix factorization, *Nature*, Vol.401, pp.788-791 (1999).
- [6] Lee, D.D. and Seung, H.S.: Algorithms for Non-negative Matrix Factorization, *Proc. NIPS'01*, pp.556-562 (2001).
- [7] Liu, W., Zheng, N. and Lu, X.: Non-negative Matrix Factorization for Visual Coding, *Proc. ICASSP'03*, pp.293-296 (2003).
- [8] 荻野広樹, 吉田哲也: トピックグラフに基づく NMF を用いた転移学習, 情報処理学会論文誌 数理モデル化と応用, Vol.4, No.3, pp.73-83 (2011).
- [9] Olshausen, B.A. and Field, D.J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images, *Nature*, Vol.381, pp.607-609 (1996).
- [10] Tibshirani, R.: Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society, Series B*, Vol.58, No.1, pp.267-288 (1996).
- [11] 上田修功, 齊藤克巳: 多重トピックテキストの確率モデル—テキストモデル研究の最前線 (1), 情報処理, Vol.45, No.2, pp.184-190 (2004).
- [12] von Luxburg, U.: A Tutorial on Spectral Clustering, *Statistics and Computing*, Vol.17, No.4, pp.395-416 (2007).
- [13] Xu, W., Liu, X. and Gong, Y.: Document clustering based on non-negative matrix factorization, *Proc. SIGIR'03*, pp.267-273 (2003).
- [14] 福高雅夫: 非線形最適化の基礎, 朝倉書店 (2001).

付 録

A.1 提案アルゴリズムの収束性

3章で提案したアルゴリズムの収束性を証明する。目的関数の値が各変数 (\mathbf{U} , \mathbf{V}) に対して単調非増加であることを示す。式 (11) は式 (14) で $\lambda_2 = 0$ とした場合に対応するため、定理 2 での \mathbf{U} に関する収束性を示す。 \mathbf{V} に関する収束性は文献 [6] を参照されたい。このため、更新式 (14) が式 (13) に対して単調非増加であることを示す。

定義 1. 以下の 2 つを満たす関数 $G(u, u^{(\tau)})$ を関数 $F(u)$ の補助関数と呼ぶ [6].

- $G(u, u^{(\tau)}) \geq F(u)$
- $G(u, u) = F(u)$

任意の補助関数に対し，補題 3 が成り立つ．

補題 3. 関数 G が関数 F の補助関数であれば F は式 (A.1) において単調非増加である [6].

$$u^{(\tau+1)} = \underset{u}{\operatorname{argmin}} G(u, u^{(\tau)}) \quad (\text{A.1})$$

$u^{(\tau)}$ は τ 回目の更新時での値を表し，式 (A.1) を通じて $u^{(\tau+1)}$ に更新される．

Proof. 関数 G は関数 F の補助関数であるため，定義 1 より以下の不等式が成り立つ．

$$F(u^{(t+1)}) \leq G(u^{(t+1)}, u^{(\tau)}) \leq G(u^{(\tau)}, u^{(\tau)}) = F(u^{(\tau)}) \quad \square$$

式 (13) が行列 \mathbf{U} の各要素 $(\mathbf{U})_{ab}$, $a \in \{1, \dots, p\}$, $b \in \{1, \dots, q\}$ の更新に対して単調非増加であることを示す．以下では任意の要素 $(\mathbf{U})_{ab}$ を u_{ab} と表記し，式 (13) で要素 $(\mathbf{U})_{ab}$ に関与する項を関数 F_{ab} と表記する．式 (13) を任意の $(\mathbf{U})_{ab}$ で微分する．

$$\begin{aligned} F'_{ab} &\triangleq \left(\frac{\partial J_2}{\partial \mathbf{U}} \right)_{ab} \\ &= (-2\mathbf{X}\mathbf{V}^T + 2\mathbf{U}\mathbf{V}\mathbf{V}^T + 2\lambda_1 \mathbf{U} \mathbf{1}_q \mathbf{1}_q^T + 2\lambda_2 \mathbf{L}\mathbf{U})_{ab} \\ \left(\frac{\partial F'_{ab}}{\partial \mathbf{U}} \right)_{ab} &= 2(\mathbf{V}\mathbf{V}^T)_{bb} + 2\lambda_1 + 2\lambda_2(\mathbf{L})_{bb} \end{aligned} \quad (\text{A.2})$$

補題 4. 以下の関数は関数 F_{ab} に対する補助関数である [1].

$$\begin{aligned} G(u, u_{ab}^{(\tau)}) &= F_{ab}(u_{ab}^{(\tau)}) + F'_{ab}(u_{ab}^{(\tau)})(u - u_{ab}^{(\tau)}) \\ &\quad + \frac{(\mathbf{U}\mathbf{V}\mathbf{V}^T)_{ab} + \lambda_1 u_{ab}^{(\tau)} + \lambda_2 (\mathbf{D}\mathbf{U})_{ab}}{u_{ab}^{(\tau)}} (u - u_{ab}^{(\tau)})^2 \end{aligned} \quad (\text{A.3})$$

Proof. 関数 G が関数 F_{ab} の補助関数であることを示す． $G(u, u) = F_{ab}(u)$ は明らか．よって，定義 1 の $G(u, u_{ab}^{(\tau)}) \geq F_{ab}(u)$ を示す． $F_{ab}(u)$ を $u^{(\tau)}$ でテイラー展開し以下を得る．

$$\begin{aligned} F_{ab}(u) &= F_{ab}(u_{ab}^{(\tau)}) + F'_{ab}(u_{ab}^{(\tau)})(u - u_{ab}^{(\tau)}) \\ &\quad + [(\mathbf{V}\mathbf{V}^T)_{bb} + \lambda_1 + \lambda_2(\mathbf{L})_{bb}](u - u_{ab}^{(\tau)})^2 \end{aligned} \quad (\text{A.4})$$

$G(u, u_{ab}^{(\tau)}) \geq F_{ab}(u)$ を示すため式 (A.3) と式 (A.4) を比較し，第 3 項の $(u - u_{ab}^{(\tau)})^2$ は非負であるため，その係数に対して以下が成り立つことを示す．

$$\begin{aligned} \frac{(\mathbf{U}\mathbf{V}\mathbf{V}^T)_{ab} + \lambda_1 u_{ab}^{(\tau)} + \lambda_2 (\mathbf{D}\mathbf{U})_{ab}}{u_{ab}^{(\tau)}} &\geq (\mathbf{V}\mathbf{V}^T)_{bb} + \lambda_1 + \lambda_2(\mathbf{L})_{bb} \geq 0 \end{aligned} \quad (\text{A.5})$$

式 (A.5) の非負性は，NMF における行列 \mathbf{U} , \mathbf{V} の非負性，提案法での λ_1 , λ_2 の非負性，および \mathbf{L} の対角要素の

非負性より成り立つ．同様に， \mathbf{U} , \mathbf{V} の非負性と次数行列 \mathbf{D} の定義から $\mathbf{D}_{bb} \geq (\mathbf{D} - \mathbf{W})_{bb}$ となるため，以下の不等式が成立する．

$$\begin{aligned} (\mathbf{U}\mathbf{V}\mathbf{V}^T)_{ab} &= \sum_{k=1}^q (\mathbf{U})_{ak}^{(\tau)} (\mathbf{V}\mathbf{V}^T)_{kb} \geq (\mathbf{U})_{ab}^{(\tau)} (\mathbf{V}\mathbf{V}^T)_{bb} \\ \lambda_2 (\mathbf{D}\mathbf{U})_{ab} &= \lambda_2 \sum_{k=1}^q (\mathbf{D})_{kb} (\mathbf{U})_{ak}^{(\tau)} \geq \lambda_2 (\mathbf{D})_{bb} (\mathbf{U})_{ab}^{(\tau)} \\ &\geq \lambda_2 (\mathbf{D} - \mathbf{W})_{bb} (\mathbf{U})_{ab}^{(\tau)} = \lambda_2 (\mathbf{L})_{bb} (\mathbf{U})_{ab}^{(\tau)} \end{aligned}$$

したがって

$$\begin{aligned} \frac{(\mathbf{U}\mathbf{V}\mathbf{V}^T)_{ab} + \lambda_1 u_{ab}^{(\tau)} + \lambda_2 (\mathbf{D}\mathbf{U})_{ab}}{u_{ab}^{(\tau)}} &\geq \frac{(\mathbf{U})_{ab}^{(\tau)} (\mathbf{V}\mathbf{V}^T)_{bb} + \lambda_1 u_{ab}^{(\tau)} + \lambda_2 (\mathbf{L})_{bb} (\mathbf{U})_{ab}^{(\tau)}}{u_{ab}^{(\tau)}} \\ &\geq (\mathbf{V}\mathbf{V}^T)_{bb} + \lambda_1 + \lambda_2 (\mathbf{L})_{bb} \end{aligned} \quad (\text{A.6})$$

上記より式 (A.5) が成り立つため，補題 4 が成り立つ． \square

式 (A.3) が補助関数であることと補題 3 より， $u_{ab}^{(t+1)}$ は以下となる．

$$\begin{aligned} u_{ab}^{(t+1)} &= u_{ab}^{(\tau)} - u_{ab}^{(\tau)} \frac{F'_{ab}(u_{ab}^{(\tau)})}{2(\mathbf{U}\mathbf{V}\mathbf{V}^T)_{ab} + 2\lambda_1 \mathbf{U} \mathbf{1}_q \mathbf{1}_q^T + 2\lambda_2 (\mathbf{D}\mathbf{U})_{ab}} \\ &= u_{ab}^{(\tau)} \frac{(\mathbf{X}\mathbf{V}^T + \lambda_2 \mathbf{W}\mathbf{U})_{ab}}{(\mathbf{U}\mathbf{V}\mathbf{V}^T + \lambda_1 \mathbf{U} \mathbf{1}_q \mathbf{1}_q^T + \lambda_2 \mathbf{D}\mathbf{U})_{ab}} \end{aligned} \quad (\text{A.7})$$

上記より，更新式 (A.7) は式 (13) に対して単調非増加であるため提案アルゴリズムは収束し，定理 1, 2 が成り立つ．



木村 圭吾 (正会員)

1989 年生．北海道大学工学部情報エレクトロニクス学科在学中．主に機械学習の研究に興味を持つ．

吉田 哲也 (正会員)

1968 年生．1991 年東京大学工学部航空工学科卒業．1997 年東京大学大学院博士課程修了．工学博士．同年大阪大学大学院基礎工学研究科助手．2001 年大阪大学産業科学研究科助手．2004 年北海道大学大学院情報科学研究科助教授．現在，同大学准教授．主に機械学習，知識獲得，データマイニング等の研究に興味を持つ．人工知能学会会員．