

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}

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。従来の研究では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。

Non-negative Matrix Factorization with Sparse Features

KEIGO KIMURA ^{†1} and TETSUYA YOSHIDA ^{†2}

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.

1. はじめに

文書中の単語の生起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)-6),9),12)}。主成分分析などの固有値解析に基づく手法と異なり、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)}や画像処理などでの有効性が報告されている^{3),4)}。また、文書クラスタリングに対する有効性も報告されている¹²⁾。

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できることを報告する。

2節では関連研究を紹介し、3節で提案法について説明し、4節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節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、 $\mathbf{1}$ は要素が全て1のベクトルを表し、正則行列 \mathbf{A} に対する逆行列を \mathbf{A}^{-1} で表す。本

^{†1} 北海道大学工学部

Faculty of Engineering, Hokkaido University

^{†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

稿では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仮定する。

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 を次元削減法とみなす場合には、行列 \mathbf{U} の列ベクトルを特徴とみなし、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。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。

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

*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。

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。

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応じ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。

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¹²⁾ 特徴同士の独立性を $\mathbf{1}_q^T \mathbf{U}^T \mathbf{U} \mathbf{1}_q$ と定式化し^{*2)}、下記の目的関数の最小化を提案する。

$$J_1 = \|\mathbf{X} - \mathbf{UV}\|^2 + \lambda_1 \mathbf{1}_q^T \mathbf{U}^T \mathbf{U} \mathbf{1}_q \quad (4)$$

ここで λ_1 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 で学習する特徴同士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 (4) の第二項は小さくなる。

3.1.2 更新式

式 (4) 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

*2 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。

図 1 アルゴリズム scNMF
Fig. 1 Algorithm scNMF

```

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 termination condition is not satisfied 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  $\mathbf{U}, \mathbf{V}$ 

```

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 (5) を定義する.

$$\mathcal{L} = \|\mathbf{X} - \mathbf{U}\mathbf{V}\|^2 + \lambda_1 \mathbf{1}_q^T \mathbf{U}^T \mathbf{U} \mathbf{1}_q + \text{tr}(\Psi \mathbf{U}^T) + \text{tr}(\Phi \mathbf{V}^T) \quad (5)$$

式 (5) を \mathbf{U}, \mathbf{V} に対して微分し, KKT(Karush-Kuhn-Tucker) 条件¹³⁾ より各要素ごとに以下が成立する.

$$(-\mathbf{X}\mathbf{V}^T + \mathbf{U}\mathbf{V}\mathbf{V}^T + \lambda_1 \mathbf{U}\mathbf{1}_q \mathbf{1}_q^T)_{ij} (\mathbf{U})_{ij} = 0 \quad (6)$$

$$(-\mathbf{X}^T \mathbf{U} + \mathbf{V}^T \mathbf{U}^T \mathbf{U})_{ij} (\mathbf{V})_{ij}^T = 0 \quad (7)$$

式 (6), (7) に基づき, 以下の更新式が得られる.

$$(\mathbf{U})_{ij} \leftarrow (\mathbf{U})_{ij} \frac{(\mathbf{X}\mathbf{V}^T)_{ij}}{(\mathbf{U}\mathbf{V}\mathbf{V}^T + \lambda_1 \mathbf{U}\mathbf{1}_q \mathbf{1}_q^T)_{ij}} \quad (8)$$

$$(\mathbf{V})_{ij} \leftarrow (\mathbf{V})_{ij} \frac{(\mathbf{U}^T \mathbf{X})_{ij}}{(\mathbf{U}^T \mathbf{U}\mathbf{V})_{ij}} \quad (9)$$

3.1.3 アルゴリズム scNMF

式 (8), (9) 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} を交互最適化を行う. 2 行目の終了条件は繰返し数や式 (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¹²⁾ での正規化を行う. 5 行目は提案法

における式 (4) での正規化条件を実現していることに注意されたい.

提案アルゴリズムについて以下の性質が成り立つ.

定理 1. 式 (4) 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.

収束性の証明は紙面の都合から省略する.

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. ここで, $\mathbf{W} \in \mathbb{R}_+^{p \times p}$ を類似度に基づいて構築した m -近傍グラフ *1) の重み行列として, 以下の目的関数の最小化を考える:

$$J_2 = \|\mathbf{X} - \mathbf{U}\mathbf{V}\|^2 + \lambda_1 \mathbf{1}_q^T \mathbf{U}^T \mathbf{U} \mathbf{1}_q + \lambda_2 \text{tr}(\mathbf{U}^T \mathbf{L} \mathbf{U}) \quad (10)$$

$\mathbf{L} = \mathbf{D} - \mathbf{W}$ はグラフの次数行列 \mathbf{D} *2) に基づくグラフラプラシアンである¹¹⁾. 式 (10) の第三項がデータ空間における属性に基づく正則化項であり, λ_2 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 節と同様に, 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 (10) を最小化する以下の更新式を導出する. 式 (10) で行列 \mathbf{V} に依存するのは第一項のみであるため, \mathbf{V} に対しては式 (9) 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:

*1 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.

*2 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.

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

$$(\mathbf{U})_{ij} \leftarrow (\mathbf{U})_{ij} \frac{(\mathbf{X}\mathbf{V}^T + \lambda_2 \mathbf{W}\mathbf{U})_{ij}}{(\mathbf{U}\mathbf{V}\mathbf{V}^T + \lambda_1 \mathbf{U}\mathbf{1}_q \mathbf{1}_q^T + \lambda_2 \mathbf{D}\mathbf{U})_{ij}} \quad (11)$$

3.2.3 アルゴリズム scGNMF

式 (10) 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 (11) 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 (10) 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。

収束性の証明は紙面の都合から省略する。

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) *1, 2) SRAA *2, 3) TREC *3 である。

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

表 2 TREC データセット

Table 2 TREC datasets (original representation)

dataset	# attr.	#class	#data
sports	126372	7	8580
reviews	126372	5	4069
klb	21839	6	2340
ohscal	11465	10	11162

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

*2 <http://www.cs.umass.edu/~mccallum/data.html>

*3 <http://glaros.dtc.umn.edu/gkhome/cluto/cluto/download>

した。各サンプルごとに porter stemmer *4 を用いて stemming を行い, MontyTagger *5 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 では既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, 各データ (ここでは文書) ごとに真のクラスタが既知である。各データセットに対して, 各データに対する真のクラスタと割り当てられたクラスタに基づいて正規化相互情報量 (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, 正規化相互情報量 (NMI) は以下で定義される。

$$NMI = \frac{I(\hat{C}; C)}{(H(\hat{C}) + H(C))/2} \quad (\in [0, 1]) \quad (12)$$

$H(\cdot)$ 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。

また, NMF とは与えられたデータ行列 \mathbf{X} を $\mathbf{U}\mathbf{V}$ 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。

$$approx. = \frac{\|\mathbf{X} - \mathbf{U}\mathbf{V}\|}{\|\mathbf{X}\|} \quad (13)$$

式 (13) の値が小さいほど行列 \mathbf{U}, \mathbf{V} による近似が良いことに対応する。

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) NMF⁶⁾, 2) WNMf¹²⁾, 3) GNMF¹⁾
- (b) スパース制約を考慮した手法: 4) SNMF⁷⁾, 5) NMFsc⁴⁾

Weighted NMF (WNMF) とは, Ncut 法¹¹⁾ 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¹²⁾。データ行列 \mathbf{X} に対するグ

*4 <http://www.tartarus.org/~martin/PorterStemmer>

*5 <http://web.media.mit.edu/~hugo/montytagger>

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 NMF⁶⁾ 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ラシアン $\mathbf{L} = \mathbf{D} - \mathbf{A}$ を用いた以下の目的関数の最小化を行う¹⁾。

$$J_2 = \|\mathbf{X} - \mathbf{UV}\|^2 + \lambda \text{tr}(\mathbf{V}\mathbf{L}\mathbf{V}^T) \quad (14)$$

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みなすことができるため、ONMF との比較を行った。

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、ひと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 *1。

4.1.5 実験パラメータ

対関係 (データ対, 特徴対, 属性対) における類似度としては文書処理で標準的なコサイン類似度を用い、図 1 の 2 行目での終了条件は繰返し数 30 とした。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 (14) 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 (??) の sp の値は 0.2 とした *2。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、予備実験のに基づき 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 のもとでの行列分解を行う。以下では特徴数 q をクラスタ数 k の 1, ..., 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 (NMI) を図 2 に示す (横軸は特徴数 q , 縦軸は NMI)。図中の凡

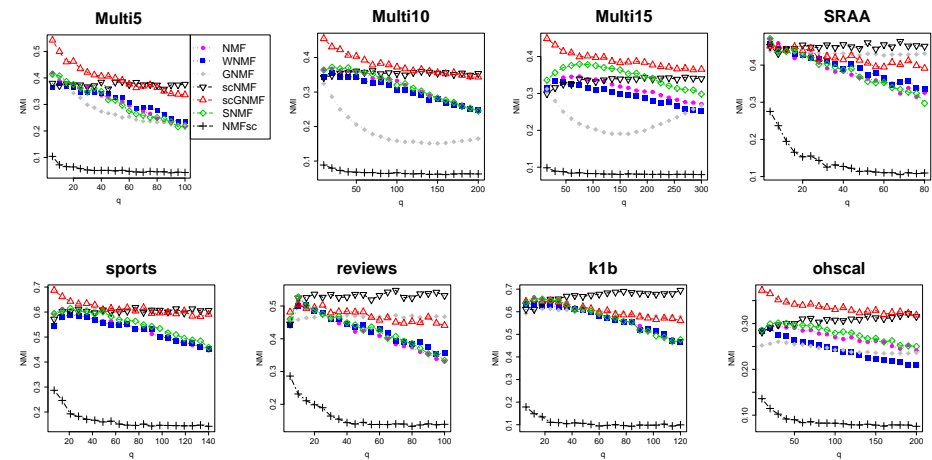


図 2 クラスタリング精度 (NMI)
Fig. 2 Clustering performance (NMI)

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 節の (a) に挙げた手法であり、丸は NMF, 四角は WNMF, 菱形は GNMF である。(b) に挙げた SNMF は菱形, NMFsc は十字で示す。

図 2 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 の増加に伴い精度が悪化した。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。

4.1.3 節の (b) に挙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。他方、NMFsc の精度は非常に低かった。

4.2.2 行列分解による近似精度

式 (13) の行列分解による近似精度を図 3 に示す (縦軸が近似精度で、小さいほど近似が良い) *3。図 3 の結果より、4.1.3 節の (b) に挙げたスパース制約を考慮した手法以外で

*1 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。

*2 スパースな行列を得るために式 (??) 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。

*3 GNMF では式 (14) の第二項が非常に大きいため式 (13) の値が 100 以上となり、非常に近似精度が悪いため図 3 では省略した。

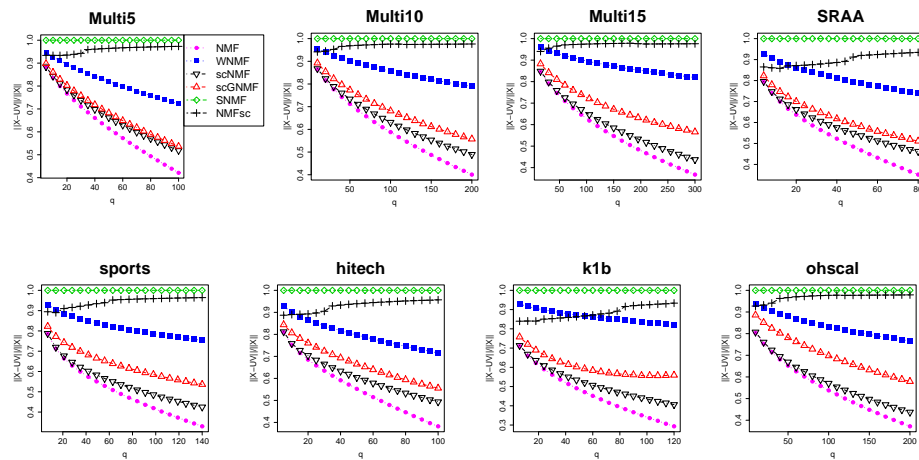


図3 データ行列の近似精度
Fig.3 Normalized quality of approximation in eq.(13)

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、scGNMF は ONMF とほぼ同程度の近似精度を示した。

他方、(b) に挙げた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とに主眼を置いているためと考えられる。

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更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) von Luxburg, U.: A Tutorial on Spectral Clustering, *Statistics and Computing*, Vol.17, No.4, pp.395–416 (2007).
- 12) Xu, W., Liu, X. and Gong, Y.: Document clustering based on non-negative matrix factorization, *Proc. SIGIR'03*, pp.267–273 (2003).
- 13) 福島雅夫: 非線形最適化の基礎, 朝倉書店 (2001).