

NMFを用いた表現学習に対する コレスキー分解を用いた補正法

吉田 哲也^{†1} 荻野 広樹^{†1}

本稿では、NMF (Non-negative Matrix Factorization) を用いて表現学習を行う際に学習される局所表現の非直交性に着目し、特徴空間における計量に対するコレスキー分解を活用することで学習した表現を補正する手法を提案する。個々のデータを局所表現の線形結合として表現する際、表現における非負性を保証する手法として NMF が注目を集めている。しかし、学習した表現を活用する際には、局所表現の非直交性に起因する問題は考慮されてこなかった。提案法では NMF がデータを局所表現の線形結合に分解することに着目し、局所表現を用いることで特徴空間における計量を推定し、計量に対するコレスキー分解から得られる上三角行列を活用して学習した表現を補正する。提案法を文書クラスタリングに適用し、提案法の有効性を確認した。

Cholesky Decomposition Rectification for Non-negative Matrix Factorization

TETSUYA YOSHIDA^{†1} and HIROKI OGINO^{†1}

We propose a method based on Cholesky decomposition for Non-negative Matrix Factorization (NMF). NMF enables to learn local representation due to its non-negative constraint. However, when utilizing NMF as a representation leaning method, the issues due to the non-orthogonality of the learned representation has not been dealt with. Since NMF learns both feature vectors and data vectors in the feature space, the proposed method 1) first estimates the metric in the feature space based on the learned feature vectors, 2) applies Cholesky decomposition on the metric and identifies the upper triangular matrix, 3) and finally utilizes the upper triangular matrix as a linear mapping for the data vectors. The proposed approach is evaluated over several real world datasets. The results indicate that it is effective and improves performance.

1. はじめに

カーネル法に代表される、与えられたデータの表現を処理の目的に適した表現に変換（写像）し、写像先の空間で処理（分類やクラスタリングなど）を行う研究が活発に行われている¹⁵⁾。分類問題に対する Support Vector Machine (SVM)¹⁵⁾、クラスタリング問題に対するスペクトルクラスタリング¹⁷⁾ など、様々な研究が行われている。本稿では、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び、データ空間におけるデータの表現をデータ行列、特徴空間におけるデータの表現を表現行列と呼ぶ。また、表現行列を学習することを表現学習と呼び、数ベクトル空間の標準基底を構成するベクトルなどのように非零の要素が少ない表現を局所表現と呼ぶ。

個々のデータを局所表現の線形結合として表現する際、表現における非負性を保証する手法として Non-negative Matrix Factorization (NMF) の研究が活発に行われている^{1),8),9),19)}。主成分分析 (PCA) などの固有値解析に基づく手法と異なり直交性が強制されない局所表現が得られ、視覚野などの研究や局所表現を活用した画像処理などでの有効性が報告されている^{5),8)}。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¹⁹⁾。

一般にスペクトル学習などの表現学習法においては、学習した表現（表現行列）を任意の学習器に適用するとして、学習器の性質を考慮しないことが多かった⁷⁾。従来法では NMF をクラスタリングに用いる際には特徴数をクラスタ数に設定する 경우가多いが^{8),19)}、NMF を表現学習ととらえて特徴数を増加させた場合に、NMF で学習する局所表現の非直交性により、通常のユークリッド空間を（暗黙に）仮定する手法を適用した場合には特徴空間の性質を反映した距離を活用できないという問題はこれまで考慮されてこなかった¹³⁾。

本稿では、非負の表現を得るために NMF を用いて表現学習を行う際、学習される局所表現の非直交性に着目し、特徴空間における計量に対するコレスキー分解⁴⁾ を用いて学習した表現を補正する手法を提案する。提案法では NMF がデータを局所表現の線形結合に分解することに着目し、局所表現を用いて特徴空間における計量を推定し、計量に対するコレスキー分解から得られる上三角行列を活用して表現行列を補正する。提案法を文書クラスタリングに適用して評価し、提案法の有効性を示す。提案法を用いて表現行列を補正することにより、NMF で学習した表現をクラスタリングなどに活用する際、適用するアルゴリズムを

^{†1} 北海道大学大学院情報科学研究科
Graduate School of Information Science and Technology, Hokkaido University

変更することなく局所表現の非直交性を補正した効果的な活用が可能になると期待される。与えられたデータに基づいて適切な計量を学習する様々な手法が提案されてきた^{10),11),18)}。しかし、計量の学習には分類学習と同様にラベル付きデータを与える必要があった。また半正定値問題として定式化されることが多く、学習に要する計算量が多かった^{18),18)}。他方、提案法ではラベル付きデータを与える必要はなく、与えられたデータのみから特徴空間における計量を推定して活用できる。さらに、コレスキー分解は高速に安定的に求めることができる。また、高次元データに対して計量の導出と次元圧縮を同時に考える必要があることが指摘されている¹¹⁾が、NMFによる次元圧縮と、圧縮した表現に対する適切な計量を導出(補正)するという意味で提案法は両者を同時に実現しているといえる。

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する。5章で提案法の有効性に関して議論し、6章でまとめと今後の展望を述べる。

2. 関連研究

2.1 次元削減

一般にカーネル法ではカーネルトリックを用いて高次元空間への非線形写像を利用することが多いが¹⁵⁾、探索的データ解析などでは文書データなどの高次元空間で表現されるデータを低次元空間に写像して見通しの良い低次元に圧縮した表現を活用することが多い。統計学では、射影行列などの線形写像を用いた手法や、主成分分析(PCA)や多次元尺度構成法などが研究されてきた⁴⁾。また、非線形写像を用いた様々な手法が提案されており、グラフラプリアンを用いた手法¹⁶⁾や局所線形性を仮定した手法¹⁴⁾などがある。

2.2 クラスタリング

大別すると階層的クラスタリングと分割的クラスタリングのアプローチがあり、従来から様々な研究が行われている⁶⁾。グラフ理論に基づいて分割的クラスタリングを行う手法が近年活発に研究されているが、このアプローチでは、類似度(あるいは非類似度)に基づきグラフの中から組合せ論的な構造を探索する。これまで、グラフ彩色に基づく手法¹²⁾やグラフカットに基づく手法(Ncut)¹⁷⁾などが提案されている。

2.3 計量学習

これまで与えられたデータ集合を用いて適切な計量を学習して活用する手法が提案されてきた^{10),11),18)}。半教師ありクラスタリングにおいて指定されたデータ対の関係から計量学習を半正定値問題として定式化する手法^{10),18)}や、訓練データにおけるクラスタ情報を用いて二次計画問題として定式化する手法¹¹⁾などが提案されている。これらは、教師情報を

活用して計量学習を行う手法といえる。

3. NMF を用いた表現学習に対するコレスキー分解を用いた補正法

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、 X_{ij} で行列 X の第 ij 要素を表す。tr は行列のトレースを表し、 X^T で X の転置を表す。

3.1 Non-negative Matrix Factorization

Non-negative Matrix Factorization (NMF)⁹⁾とは、与えられた p 次元空間における n 個のデータに対する非負のデータ行列 $X = [x_1, \dots, x_n] \in \mathbb{R}_+^{p \times n}$ に対して^{*1}、指定された特徴数 q のもとで非負の行列 $U = [u_1, \dots, u_q] \in \mathbb{R}_+^{p \times q}$ 、 $V = [v_1, \dots, v_n] \in \mathbb{R}_+^{q \times n}$ の積により

$$X \simeq UV \quad (1)$$

と近似して分解する行列 U 、 V を求める手法である。各データ x_i は u_1, \dots, u_q の線形結合として近似して表現されることになる。このため、NMF は与えられたデータ行列 X に対し、行列 U の列ベクトルで張られる特徴空間における表現 V を求める手法と見なせる。 U 、 V は式 (1) での近似の良さに対応する以下の目的関数を最小化により求める。

$$J_1 = \|X - UV\|^2 \quad (2)$$

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⁹⁾、本稿ではフロベニウスノルムを用いる場合を扱う。各データは特徴空間において q 個のベクトル u_1, \dots, u_q の線形結合として近似的に表現されるため、次元削減法と見なすことができる。PCA などの固有値解析に基づく手法と比較して、NMF では次元削減後の各要素が非負であり、また、通常のユークリッド空間のもとの各特徴ベクトルの直交性は強制されないという特徴がある。

NMF のアルゴリズムを図 1 に示す⁹⁾。1 行目で初期化し、3, 4 行目で U 、 V を交互最適化を行う。終了条件としては、繰返し数の上限や式 (2) の値が十分小さい場合とすることが多い。一般に式 (1) を満たす U 、 V は無数に存在するため、解の一意性の保証のために 2 行目の繰返し中に U に対する正規化を用いる手法が提案されている¹⁹⁾。

3.2 NMF を用いたクラスタリング

NMF は視覚野などにおける局所特徴量に対する研究で提案されてきたが⁸⁾、高次元スパースな表現を持つ文書クラスタリングでの有効性も示されている^{3),19)}。NMF を用いた文書ク

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

```

NMF(X, q)
Require: X ∈ ℝp×n: データ行列
Require: q ∈ ℕ: 特徴数
1: U, V を初期化
2: while 終了条件が満たされない限り do
3:   Uij := Uij  $\frac{(XV^T)_{ij}}{(UVV^T)_{ij}}$ 
4:   Vij := Vij  $\frac{(U^T X)_{ij}}{(U^T UV)_{ij}}$ 
5: end while
6: return U, V

```

図 1 Non-negative Matrix Factorization (NMF)
Fig. 1 Non-negative Matrix Factorization (NMF).

ラスタリングでは、一般には特徴数 q をクラスタ数 k に設定し、各データ x ごとに構築した表現 v において最大の値を持つクラスタ c に割り当てる。

$$c = \underset{c}{\operatorname{argmax}} v_c \quad (3)$$

ここで v_c はデータ x に対する表現 v における第 c 要素の値を表す。

3.3 NMF を用いた表現学習

2.1 節で述べた次元削減法は、与えられたデータ空間から扱う問題に適した空間における表現を求めるという意味で表現学習と見なせる。カーネル法では高次元空間への非線形写像を実現するカーネル関数を目的に合わせて事前に設計することが多く¹⁵⁾、グラフラプリアンを用いてデータからカーネル行列を求め、カーネル行列の固有ベクトルを用いる手法も提案されている⁷⁾。

NMF を用いた表現学習として、画像処理における局所表現を活用した画像検索での有効性が示されている。また、NMF を用いて扱う対象領域に一般的な特徴をまず学習しておき、後に様々な用途に活用する自己教示学習も提案されている¹³⁾。しかし、画像データに対しては効果的であるが、文書データに対しては効果が少ないことが報告されている¹³⁾。

NMF を次元削減を行う表現学習法ととらえた場合、学習した表現行列 V に対して SVM などの分類学習法や kmeans などのクラスタリング法を適用することになる。一般には kmeans 法などのように特徴空間を構成する特徴どうしの直交性を（暗黙に）仮定する手法を適用することが多いが⁷⁾、NMF により学習した表現（表現行列）を活用する際には、NMF で学習する局所表現の非直交性により特徴空間の性質を反映した距離を活用できないという問

題は考慮されてこなかった¹³⁾。

3.4 コレスキー分解を用いた補正法

上記の問題は、NMF で学習した表現行列 V を活用する際、特徴空間におけるデータ対 (v_i, v_j) に対する二乗距離を特徴空間が通常の単純なユークリッド空間と仮定して $(v_i - v_j)^T(v_i - v_j)$ として扱うことに起因する。しかし、NMF では非負性の制約により一般には u_1, \dots, u_q の直交性は成立しないため、学習した表現行列 V に対して上記のような二乗距離を使用することは適切ではない。特徴空間の非直交性を反映する計量 M が分かっていたら、 M を用いてデータ対 (v_i, v_j) に対する二乗距離を

$$(v_i - v_j)^T M (v_i - v_j) \quad (4)$$

として求めることができる。これは一般に画像などの分類問題でのマハラノビス距離である。

本稿では、NMF がデータ行列 X を特徴空間を構成する局所表現 U と、特徴空間における表現行列 V に分解することに着目し、1) U を用いて特徴空間における計量 M を推定する、2) M を用いて学習した表現 V を補正する、3) 補正した表現をユークリッド空間などの直交性を仮定するアルゴリズムを適用する、という手法を提案する。

上記 3) は、文献 7) と同様に補正後の表現に対して既存の手法を適用することに対応するため、以下では 1), 2) について述べる。なお、提案法により補正した表現を用いることにより、上記 3) で適用するアルゴリズムを変更することなくデータ処理が可能となる。

3.4.1 NMF における計量の推定

データ行列 X を q 個のベクトル u_1, \dots, u_q で張られる特徴空間で近似して表現する際、NMF では特徴空間を構成する各 u に対してもとのデータ空間における明示的な表現が与えられていることに着目し、3.1 節の末尾に述べたように図 1 に示すアルゴリズム中で各 u に対して $u^T u = 1$ という正規化を行い¹⁹⁾、 U のグラム行列 $U^T U$ により特徴空間における計量 M を推定する。

$$M = U^T U, \quad \text{s.t. } u_l^T u_l = 1, \quad \forall l = 1, \dots, q \quad (5)$$

これは、グラフラプリアンを用いる手法においてデータ行列 X のグラム行列 $X^T X$ をカーネルとして用いることに対応する。ただし、他の計量学習法と異なり、提案法では NMF の性質を用いることにより M の推定に教師情報を与える必要がないことに注意されたい。また、式 (1) は各データを u_1, \dots, u_q で張られる特徴空間で近似して表現するという意味で正規化した U のグラム行列を計量として用いることは自然であると考えられる。

3.4.2 コレスキー分解を用いた補正

式 (5) により計量 M を推定するため、推定した M は対称非負定値であることが保証さ

```

CNMF( $X$ ,  $algNMF$ ,  $q$ ,  $parameters$ )
Require:  $\mathbf{X} \in \mathbb{R}_+^{p \times n}$  //data matrix
Require:  $q$ ; //the number of features
Require:  $algNMF$ ; //the utilized NMF algorithm
Require:  $pars$ ; //other parameters in  $algNMF$ 
1:  $\mathbf{U}, \mathbf{V} := \text{run } algNMF \text{ on } \mathbf{X}$  with  $q$  (and  $pars$ ) s.t.  $\mathbf{u}_l^T \mathbf{u}_l = 1, \forall l = 1, \dots, q$ 
2:  $\mathbf{M} := \mathbf{U}^T \mathbf{U}$ 
3:  $\mathbf{T} := \text{cholesky decomposition of } \mathbf{M}$  s.t.  $\mathbf{M} = \mathbf{T}^T \mathbf{T}$ 
4: return  $\mathbf{U}, \mathbf{TV}$ 

```

図 2 NMF に対するコレスキー分解を用いた補正法 (CNMF)
Fig. 2 Cholesky Decomposition Rectification for NMF (CNMF).

れる。このため、コレスキー分解を用いることにより、 \mathbf{M} を上三角行列 \mathbf{T} を用いて一意に分解できることが保証される⁴⁾。

$$\mathbf{M} = \mathbf{T}^T \mathbf{T} \quad (6)$$

式 (6) を式 (4) に代入することにより、コレスキー分解で求めた上三角行列 \mathbf{T} を用いて

$$\mathbf{V} \rightarrow \mathbf{TV} \quad (7)$$

と線形写像することにより、補正した表現 \mathbf{TV} を構築する。提案法により補正した表現では任意のデータ対 $(\mathbf{Tv}_i, \mathbf{Tv}_j)$ の二乗距離は

$$(\mathbf{Tv}_i - \mathbf{Tv}_j)^T (\mathbf{Tv}_i - \mathbf{Tv}_j) = (\mathbf{v}_i - \mathbf{v}_j)^T \mathbf{T}^T \mathbf{T} (\mathbf{v}_i - \mathbf{v}_j) \quad (8)$$

$$= (\mathbf{v}_i - \mathbf{v}_j)^T \mathbf{M} (\mathbf{v}_i - \mathbf{v}_j) \quad (9)$$

となり、式 (4) と同様に特徴空間の計量を反映した二乗距離となることが保証される。

なお、一般に対称非負定値行列 \mathbf{A} は $\mathbf{A} = \mathbf{P}^T \mathbf{P}$ の形に分解可能であるが⁴⁾、提案法では分解の一貫性、高速性、安定性のためにコレスキー分解を用いている。

3.5 アルゴリズム

提案するアルゴリズムを図 2 に示す。1 行目で指定された NMF アルゴリズムを用いて行列 \mathbf{U}, \mathbf{V} を求める。2 行目で計量 \mathbf{M} を構築し、3 行目で \mathbf{M} に対するコレスキー分解により上三角行列 \mathbf{T} を求め、 \mathbf{U} と補正した表現行列 \mathbf{TV} を返す。

4. 評価

4.1 実験設定

4.1.1 対象データ

提案法を 20 ニュースグループ (以下、20NG と表記)¹⁾、TREC データセット²⁾のクラ

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.	#classes	#data
hitech	126,372	6	2,301
reviews	126,372	5	4,069
sports	126,372	7	8,580
fbis	2,000	17	2,463
ohscal	11,465	10	11,162
k1b	21,839	6	2,340

スタリングに対して評価した。これらは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である。一般に文書に含まれる単語数は膨大であり、高次元スパースなデータのクラスタリングに対応する。

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^{*3)}を用いて stemming を行い、MontyTagger^{*4)}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。TREC データセットに対しては表 2 に示す 6 つのデータセットを使用した。20NG と同様な処理により^{*5)}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 データ

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

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

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

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

*5 fbis のみ各クラスタごとに 35 個ずつ文書を非復元抽出した。

セット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、各データ（ここでは文書）ごとに真のクラスが既知である。各データセットに対して、各データに対する真のクラスと割り当てられたクラスに基づいて正規化相互情報量（ NMI ）を評価した。真のクラスと割り当てられたクラス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。

$$NMI = \frac{I(\hat{C}; C)}{(H(\hat{C}) + H(C))/2} \quad (\in [0, 1]) \quad (10)$$

$H(\cdot)$ 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でのデータ割当てに合致することを示すため、クラス割当ての精度に対応する。

4.1.3 比較手法

代表的な NMF の手法である、1) NMF⁹⁾、2) WNMf¹⁹⁾、3) GNMF¹⁾ に対し、提案法による補正の効果を確認した。比較手法はすべて分割のクラスタリングを行うものであるため、クラス数 k は与えられると仮定した。また、高次元スパースデータを扱う文書クラスタリングに対する標準的な手法として skmeans²⁾ をベースラインとして評価した。

Weighted NMF (WNMF) とは、Ncut 法¹⁷⁾ 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¹⁹⁾。データ行列 X に対するグラム行列 $X^T X$ を重み行列 W とするグラフを考え、 W に対する次数行列^{*1} D の逆行列を Γ としてデータ行列 X を $X\Gamma^{1/2}$ に変換し、変換した $X\Gamma^{1/2}$ に対して図 1 に示した NMF を適用して表現行列を求める。

Graph regularized NMF (GNMF) とは、データ空間における m -近傍グラフを用い、グラフラプリアンを正則化項として用いる手法である¹⁾。 m -近傍グラフの隣接行列 A に対する次数行列 D を用いて、 m -近傍グラフに対するグラフラプリアン $L = D - A$ を用いた以下の目的関数の最小化を行い、表現行列 V を求める。

$$J_2 = \|X - UV\|^2 + \lambda \text{tr}(VLV^T) \quad (11)$$

ここで λ は第 2 項に対する正則化パラメータである。

skmeans とは、各データを $x^T x = 1$ と正規化することにより超球面上に写像し、超球面上でのコサイン類似度を用いる kmeans 法である²⁾。

4.1.4 実験パラメータ

図 1, 図 2 に示すように、NMF における主要パラメータは特徴数 q である。このため、 q をクラス数 k の 1, ..., 20 倍まで変えて実験した。図 1 の 2 行目での終了条件は繰返し数 30 とした。

データ空間における類似度としては、文書処理で標準的に用いられるコサイン類似度を用いた。GNMF に対しては各データごとに類似度が上位 10 個の近傍データを用いて m -近傍グラフを構築し ($m = 10$)、式 (11) での λ は文献 1) で推奨された値 ($\lambda = 100$) とした。

4.1.5 実験手順

各手法で構築した表現行列 V 、および提案した式 (7) で変換した行列 TV に対して、ユークリッド空間を仮定する標準的なクラスタリング法として kmeans 法および 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 U, V ）は初期値に依存する。このため、同一のデータ行列に対して図 1 の 1 行目の初期化をランダムに 10 回変えて実験した。また、kmeans 法および skmeans 法によるクラス割当ては初期割当ての影響を受けるため、生成した表現に対して初期割当てをランダムに 10 回変えて実行した。

4.2 実データに対する評価

本節での図は各データセットごとに対する 10 サンプルの平均^{*2}であり、横軸は特徴数 q 、縦軸は式 (10) の NMI に対応する。図中の凡例で実線は NMF、点線は WNMf、破線は GNMF である。なお、+c は各手法による表現行列 V に式 (7) で提案した TV に対する結果である。また、比較手法に対して従来法 (3.2 節における式 (3)) による各手法の結果^{*3}、および skmeans の結果を表 3 に示す。以下で示すように、提案法を用いて補正した表現に skmeans を適用した場合（特に WNMf+c）は表 3 に示す従来法と同等以上の性能を示した。

4.2.1 20 ニュースグループ

表 1 に示した 20NG に対する結果を図 3 に示す。図 3 より、NMF により学習した表現 V に対して通常のユークリッド距離を用いる kmeans 法およびコサイン類似度を用いる skmeans 法を適用する際、提案法を用いることで精度が向上していることが分かる。この結果から提案法は効果的であるといえる。特に、skmeans 法に対する向上率が高く（図 3 の下段）、表 3 と比較しても精度（ NMI ）は高かった。また、WNMF に対して提案法を用い

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

*2 サンプルごとに 100 回試行したため、各特徴数ごとに計 1,000 回の試行の平均値を示す。

*3 各データセットごと 10 サンプル作成したため、計 100 回の試行の平均値を示す。

表 3 式 (3) を用いた場合および skmeans の結果 (NMI)
Table 3 Results with Eq. (3) and skmeans (NMI).

	Multi5	Multi10	Multi15	hitech	reviews	sports	fbis	ohscal	k1b
NMF	0.266	0.215	0.247	0.285	0.321	0.499	0.357	0.252	0.523
WNMF	0.310	0.295	0.299	0.307	0.341	0.554	0.505	0.276	0.588
GNMF	0.124	0.065	0.068	0.067	0.260	0.387	0.364	0.076	0.227
skmeans	0.371	0.335	0.330	0.270	0.468	0.575	0.562	0.261	0.602

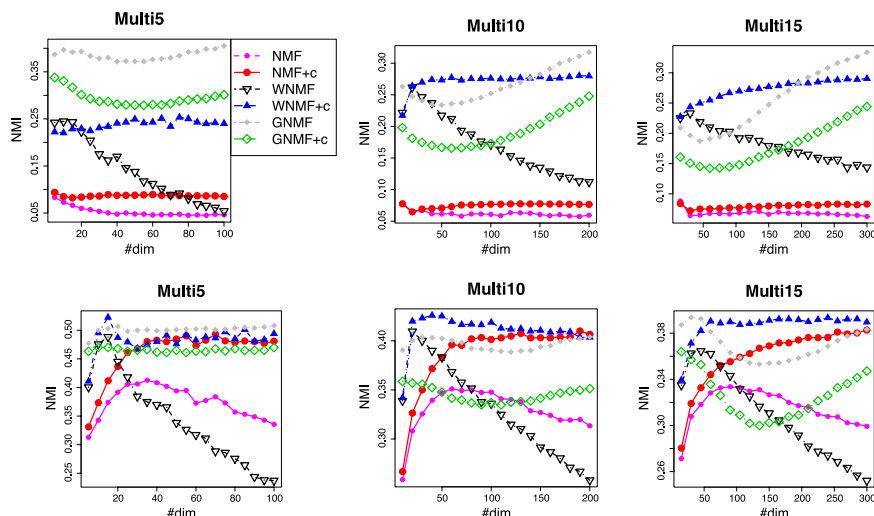


図 3 20 Newsgroup に対する結果 (NMI) (上段: kmeans, 下段: skmeans)

Fig. 3 Results on 20 Newsgroup datasets (NMI) (upper: kmeans, lower: skmeans).

た場合 (青色点線の WNMF+c) は GNMF と比較しても同等か上回る精度を示した。

他方, GNMF には提案法は効果がないことが分かる (逆に悪化した)。この理由として, GNMF の目的関数である式 (11) では, 通常の NMF での目的関数に対応する第 1 項に加えて第 2 項を正則化項として追加し, 文献 1) に従って大きな正則化係数 (λ) を用いている。式 (11) の値に対する第 2 項からの影響が非常に大きいため, NMF における式 (1) の近似関係が成り立たず, 構築される表現 V はデータ行列 X の U に基づく線形近似とはかけ離れた表現になっていると考えられる。このため, 3.4 節で述べた枠組みが成立しないために

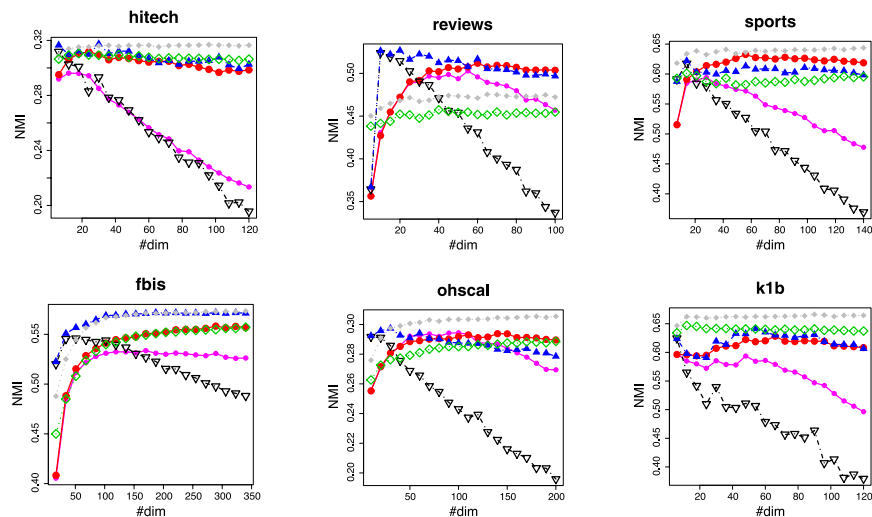


図 4 TREC データセットに対する結果 (NMI)

Fig. 4 Results on TREC datasets (NMI).

提案法の効果がなかったと考えられる。GNMF は精度が高い手法であるため, GNMF で学習される表現行列 V の補正は興味深い, 今後の課題である。

特徴数 q の増加にともない NMF, WNMF の精度は急激に悪化した, 提案法の補正を用いた場合 (NMF+c, WNMF+c) には非常にロバストであった。特に, WNMF は特徴数 q の増加にともない急激に精度が悪化した, 提案法を用いた WNMF+c は q の増加に対してロバストな結果を示し, 精度も高かった。表現学習, 特に自己教示学習では多くの局所表現を学習して活用することが重要となるため¹³⁾, この観点からも提案法は効果的であると考えられる。

4.2.2 TREC データセット

表 2 に示した TREC データセットに対する結果を図 4 に示す。NMF による表現行列 V に対する結果としては, 20NG と同様に skmeans 法が kmeans 法を大きく上回り, 両者の関係もほぼ同様な傾向を示したため, 図 4 には skmeans 法の結果のみを示す。

TREC データセットでも, 20NG に対する結果とほぼ同様な結果となった。提案法はクラスタリング手法としても (k1b を除いては) GNMF とほぼ同程度の性能を示し (特に

WNMF+c), また特徴数 q に対して非常にロバストであった. 他方, 20NG の場合と比較して, TREC データセットでは NMF+c と WNMF+c とで精度の違いはそれほど大きくなかった.

5. 議 論

5.1 提案法の有効性

本稿では, 固有値解析を用いるスペクトル学習法⁷⁾ などと異なり, NMF を用いた表現学習における非直交性に起因する問題を指摘し, コレスキー分解を用いた補正法を提案し, 4章の実験により提案法の有効性を確認した. これは, 式 (9) で示したように提案法では特徴空間の計量を反映した距離を活用できるためと考えられる. 特に, NMF を用いた表現学習においては特徴数 q が増えるほど非直交性に起因する問題が大きくなると考えられるが, 提案法は q に対して非常にロバストであることを確認した. また, 高次元スパースな表現を持つ文書クラスタリングに対しても GNMF¹⁾ とほぼ同程度の性能を示したため, 提案法はクラスタリングに対しても有効であると考えられる.

他方, データ空間における m -近傍グラフを用いた正則化項を追加した GNMF に対しては, 正則化項の影響が非常に大きいため提案法の枠組みが成立せず, 効果的ではなかった. 正則化項を活用する手法は近年活発に研究されているため^{5),13)}, 正則化項を追加した場合の扱いは興味深い課題として残されている.

5.2 計量学習法としての有効性

適切な計量を学習して活用する様々な手法が提案されてきた^{10),11),18)}. 従来の手法では, 式 (4) に基づいて計量 M に対する $M^{1/2}$ を線形写像として用いる変換を考え, $M^{1/2}$ で写像された空間 (特徴空間) において, 高い分散を抑え, 低い分散を拡大するような M の学習を行うことが多い. しかし, 適切な計量を学習するためには分類学習と同様にラベル付きデータを与える必要があった. 他方, 提案法ではラベル付きデータを与える必要はなく, 与えられたデータのみから学習した表現に対する計量を推定し, 推定した計量に基づいて表現を補正して活用することが可能である.

従来法では半正定値問題や二次計画問題として定式化されることが多く, 計量学習に要する計算量が大きかった. 図 1 に示す交互最適化を行う NMF は行列演算のみであるため高速であり, 式 (5) による計量 M の推定も内積計算であるため高速に実行できる. さらに, p 次元のデータ空間を特徴数 q の特徴空間に次元縮約するため, 提案法での特徴空間に対する計量 M のサイズはデータ空間に対する計量と比較して一般には非常に小さくなり, コレ

スキー分解により上三角行列 T を高速に安定的に求めることが可能である. 二次計画問題として定式化した手法¹¹⁾ は半正定値問題として定式化する手法¹⁸⁾ などに比べれば高速ではあるが, skmeans のような超球面上での手法には意味のある計量を与えないことが報告されている. 他方, 本稿での提案法では, skmeans を用いた場合にも効果があった.

また, 高次元データに対して計量の導出と次元圧縮を同時に考える必要があることが指摘されている¹¹⁾. 提案法は, NMF による次元圧縮と, 圧縮した表現に対する適切な計量を導出 (補正) するという意味で両者を同時に実現しているといえる.

6. おわりに

本稿では, NMF (Non-negative Matrix Factorization) を用いて表現学習を行う際, 学習される局所表現の非直交性に着目し, 特徴空間における計量に対するコレスキー分解⁴⁾ を用いて学習した表現を補正する手法を提案した. 提案法では NMF がデータ行列を局所表現とその表現行列に分解することに着目し, 局所表現を用いて特徴空間における計量を推定し, 計量に対してコレスキー分解から得られる上三角行列を活用して表現行列を補正する. 提案法は非常にシンプルではあるが, 提案する補正法を用いて表現行列を変換することにより, NMF で学習した表現をクラスタリングなどに活用する際, 適用するアルゴリズムを変更することなく局所表現の非直交性を補正した効果的な活用が可能になると期待される.

提案法を 20 ニュースグループや TREC データセットなどの高次元スパースな表現を持つ実データに対して評価し, クラスタリングの際に NMF で学習した表現を補正して効果的に活用できることを確認した. 特に, NMF を用いた表現学習においては特徴数が増えるほど非直交性に起因する問題が大きくなると考えられるが, 提案法は特徴数に対して非常にロバストであることを確認した. 今後は画像などの他の実データに対しても評価して, さらなる改良を行うとともに, 正則化項を追加した場合への拡張に取り組む予定である.

謝辞 有益なご指摘を賜りました査読者の方々に深く謝意を表します.

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) Dhillon, J. and Modha, D.: Concept decompositions for large sparse text data using clustering, *Machine Learning*, Vol.42, pp.143–175 (2001).
- 3) Ding, C., Li, T., Peng, W. and Park, H.: Orthogonal Nonnegative Matrix Tri-factorizations for Clustering, *Proc. KDD '06*, pp.126–135 (2006).

- 4) Harville, D.A.: *Matrix Algebra From a Statistician's Perspective*, Springer (2008).
- 5) Hoyer, P.O.: Non-negative Matrix Factorization with Sparseness Constraints, *Journal of Machine Learning Research*, Vol.5, pp.1457–1469 (2004).
- 6) Jain, A., Murty, M. and Flynn, P.J.: Data Clustering: A Review, *ACM Computing Surveys*, Vol.31, pp.264–323 (1999).
- 7) Kamvar, S.D., Klein, D. and Manning, C.D.: Spectral learning, *Proc. IJCAI'03*, pp.561–566 (2003).
- 8) Lee, D.D. and Seung, H.S.: Learning the parts of objects by non-negative matrix factorization, *Nature*, Vol.401, pp.788–791 (1999).
- 9) Lee, D.D. and Seung, H.S.: Algorithms for Non-negative Matrix Factorization, *Proc. Neural Information Processing Systems (NIPS)*, pp.556–562 (2001).
- 10) Li, Z., Liu, J. and Tang, X.: Pairwise constraint propagation by semidefinite programming for semi-supervised classification, *Proc. ICML'08*, pp.576–583 (2008).
- 11) 持橋大地, 菊井玄一郎, 北 研二: 言語表現のベクトル空間モデルにおける最適な計量距離, *電子情報通信学会論文誌*, Vol.J88-D-II, No.4, pp.747–756 (2005).
- 12) Ogino, H. and Yoshida, T.: Toward improving re-coloring based clustering with graph b-coloring, *Proc. PRICAI-2010*, pp.206–218 (2010).
- 13) Raina, R., Battle, A., Lee, H., Packer, B. and Ng, A.: Self-taught learning: Transfer learning from unlabeled data, *Proc. ICML'07*, pp.759–766 (2007).
- 14) Roweis, S. and Saul, L.: Nonlinear Dimensionality Reduction by Locally Linear Embedding, *Science*, Vol.290, No.22, pp.2323–2326 (2000).
- 15) Shawe-Taylor, J. and Cristianini, N.: *Kernel Methods for Pattern Analysis*, Cambridge University Press (2004).
- 16) Tenenbaum, J., de Silva, J. and Langford, J.: A Global Geometric Framework for Nonlinear Dimensionality Reduction, *Science*, Vol.290, No.22, pp.2319–2323 (2000).
- 17) von Luxburg, U.: A Tutorial on Spectral Clustering, *Statistics and Computing*,

Vol.17, No.4, pp.395–416 (2007).

- 18) Xing, E.P., Ng, A.Y., Jordan, M.I. and Russell, S.: Distance metric learning, with application to clustering with side-information, *NIPS 15*, pp.505–512 (2003).
- 19) Xu, W., Liu, X. and Gong, Y.: Document clustering based on non-negative matrix factorization, *Proc. SIGIR'03*, pp.267–273 (2003).

(平成 23 年 2 月 5 日受付)

(平成 23 年 3 月 13 日再受付)

(平成 23 年 3 月 25 日採録)



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所助手。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。



荻野 広樹 (正会員)

1986 年生。2008 年北海道大学工学部情報エレクトロニクス学科卒業。2010 年北海道大学大学院情報科学研究科修士課程修了。現在、NTT データ勤務。主に機械学習に興味を持つ。