F-004

平行移動不変な非負値行列因子分解とその分析

Shift-Invariant Non-negative Matrix Factrization and its Empirical Evaluation

鈴木 慶介 † 今井 英幸 † 張 若霓 † 瀧川 一学 †‡ 湊 真一 †
Keisuke Suzuki Hideyuki Imai Ruoni Zhang Ichigaku Takigawa Shin-ichi Minato

1. はじめに

NMF(非負値行列因子分解，Non-negative Matrix Factorizaton) は Lee と Seung[1] により，要素が非負値のデータ行列を二つの非負値行列の積に分解する手法として提案され，様々な応用で広く利用されている．NMF は非負値制約により，分解前のデータを分解後の行列で “加法的” に表現できるため，分解結果がより直感的に解釈し易くなる．この特性により，ソフトクラスタリングや特徴抽出の手法としてよく用いられてきた．しかし，NMF はデータの平行移動に対して不変ではないことが分かっている．これはデータの平行移動によってクラスタリング結果が変わりうることを意味する．例えば，k-means 法など，通常のクラスタリング手法では，データの相対的な分布形状により結果が定まるため，平行移動に対して結果は不変である．一方，NMF ではデータを分解行列の列ベクトルの錐結合で表現することに対応するため，原点の取り方に依存しない相対的な分布形状の表現とならず，平行移動不変性が成り立たない．本研究では，錐結合による分解を凸結合による分解とすることで，平行移動に対して不変な結果を出力する NMF を提案する．

2. NMF(非負値行列因子分解)

データとして，非負の要素を持つ p 次元のベクトル $\mathbf{x} = (x_1, x_2, ..., x_p)^T, x_i \geqslant 0$ が n 個与えられ，その全体をデータ行列 $\boldsymbol{X} = [\mathbf{x}_1, ..., \mathbf{x}_n]$ で表す．また，行列 $\boldsymbol{X}$ の要素が全て非負値であることを $\boldsymbol{X} \geq 0$ と表記する．NMF は行列 $\boldsymbol{X} \in \mathbb{R}^{p\times n}, \boldsymbol{X} \geq 0$ を

$$\boldsymbol{X} = \boldsymbol{F}\boldsymbol{G}^T, \quad \boldsymbol{F} \geq 0, \boldsymbol{G} \geq 0$$

となるような二つの非負値行列 $\boldsymbol{F} \in \mathbb{R}^{p\times k}$ と $\boldsymbol{G} \in \mathbb{R}^{n\times k}$ の積に分解しようとする計算に対応する．ここで，k はデータ全体を表現するのに必要な代表ベクトルの数を指し，プログラム実行者があらかじめ決めておく．クラスタリングとして見れば，p 次元のデータベクトル n 個を，p 次元の代表ベクトル k 個の錐結合で表現することに対応する．従って，k はデータの次元 p，データの個数 n よりも小さい値とする．このとき，一般には等号は成り立たず上記のような分解は得られない．そこで，NMF では $\boldsymbol{X}$ と $\boldsymbol{F}\boldsymbol{G}^T$ の二乗誤差を目的関数 $J(\boldsymbol{F}, \boldsymbol{G})$ とし，

$$J(\boldsymbol{F}, \boldsymbol{G}) = \|\boldsymbol{X} - \boldsymbol{F}\boldsymbol{G}^T\|_F^2$$

を最小化する二つの行列に分解する．ここで $\|\cdot\|_F$ は行列のフロベニウスノルムである．なお，この二乗誤差の最小化は x_{wi}が, $\sum_{t=1}^{k} f_{wt}g_{ti}$ を平均とした正規分布により生成されると仮定した場合の統計的推定問題に対応する．$\boldsymbol{F}, \boldsymbol{G}$ の $\boldsymbol{X}$ からの乖離度を表現する基準に応じて異なる最適化問題に帰着するが，本研究では上記の $J(\boldsymbol{F}, \boldsymbol{G})$ を対象とする．

NMF では，観測データの各々 $\boldsymbol{x}_i$ を

$$\boldsymbol{x}_i \approx \sum_{t=1}^{k} \boldsymbol{f}_t g_{ti} \quad g_{ti} \geq 0, i \in \{1, ..., n\} \qquad (1)$$

という形で，行列 $\boldsymbol{G}$ の要素 g_{ti} を係数，$\boldsymbol{F}$ の列ベクトルを基底ベクトルとした $\boldsymbol{f}_t$ の錐結合で最小二乗近似しようとしていると解釈できる．従って，NMF による分解は元の観測データを最も良く近似する錐包を張る k 個の基底ベクトルを探索する作業と言える．また，各基底ベクトル $\boldsymbol{f}_t$ の係数 g_{ti} の比をソフトクラスタリングの結果として解釈することが可能である．目的関数 $J(\boldsymbol{F}, \boldsymbol{G})$ は両変数に対し同時に凸ではないので解析的に解を定めることは出来ないが，以下の二つの更新式の反復で求められることが知られている [2].

$$\boldsymbol{F} \leftarrow \boldsymbol{F} \odot \frac{\boldsymbol{X}\boldsymbol{G}}{\boldsymbol{F}\boldsymbol{G}^T\boldsymbol{G}}, \quad \boldsymbol{G} \leftarrow \boldsymbol{G} \odot \frac{\boldsymbol{X}^T\boldsymbol{F}}{\boldsymbol{G}\boldsymbol{F}^T\boldsymbol{F}}$$

ここで，$\odot$ は要素同士の乗算を行う演算子である．また行列同士の除算に関しても，本文では要素同士の除算とする．この更新について目的関数は単調減少し局所最適解に収束するが，必ずしも大域最適解には収束しない．

3. Shift-Invariant NMF

係数 g_{ti} をデータ点 $\boldsymbol{x}_i$ の代表ベクトル $\boldsymbol{f}_t$ への寄与度と解釈し，ソフトクラスタリングを行うことを考える．データ点 $\boldsymbol{x}_i$ が平行移動 $\boldsymbol{a}$ を受け $\boldsymbol{x}_i + \boldsymbol{a}$ となるとき，(1) の右辺は

$$\textstyle\sum_{t=1}^{k}(\boldsymbol{f}_t + \boldsymbol{a})g_{ti} = \sum_{t=1}^{k} \boldsymbol{f}_t g_{ti} + (\sum_{t=1}^{k} g_{ti})\boldsymbol{a} \qquad (2)$$

となり，第二項分の乖離が生じるため，実際には基底ベクトルとして $\boldsymbol{f}_t + \boldsymbol{a}$ は選択されず，クラスタへの寄与度 g_{ti} も当然変化してしまう．従って，一般に NMF には平行移動による不変性がないことが分かる．

しかし，クラスタリングにおいては，移動前，移動後に関わらず観測データに対し基底との位置的な関係が不変であることが望ましい．つまり，データ行列 $\boldsymbol{X}$ に対し，列ベクトル $\boldsymbol{a} = (a_1, ..., a_p)^T$ だけ平行移動を行ったデータを NMF で分解したとき，平行移動前のデータが $\boldsymbol{X} \approx \boldsymbol{F}\boldsymbol{G}^T$ と分解されているとすると，

$$\boldsymbol{X} + \boldsymbol{a}\mathbf{1}_n^T \approx \tilde{\boldsymbol{F}}\tilde{\boldsymbol{G}}^T = (\boldsymbol{F} + \boldsymbol{a}\mathbf{1}_n^T)\boldsymbol{G}^T$$

の様に，基底行列 $\boldsymbol{F}$ に含まれる基底ベクトル $\boldsymbol{f}_i$ を同じくベクトル $\boldsymbol{a}$ だけ平行移動させた行列と，$\boldsymbol{G}$ による積で表現されなければならない．ここで $\mathbf{1}_n = (1, 1, ..., 1)^T$ であり，1 を n 個要素に持つような列ベクトルとして定義する．式 (2) からも示唆される通り，以上の条件を満たす Shift-Invariant NMF (SINMF) は以下の形で定義できる．

$$\min \boldsymbol{J}(\boldsymbol{F}, \boldsymbol{G}) \text{ subject to } \boldsymbol{F} \geq 0, \boldsymbol{G} \geq 0, \textstyle\sum_{t=1}^{k} g_{ti} = 1$$

ここで g_{ti} は $\boldsymbol{G}$ の t 行 i 列目の要素を示す．以上の制約により，データ行列の非負値性を崩さない範囲での平行移動 $\boldsymbol{a}, \boldsymbol{X} + \boldsymbol{a}\mathbf{1}_n^T \geq 0$ の下で，SINMF は目的関数が不変である．

両手法の差として，基底による空間の表現方法の違いが挙げられる．通常の NMF は基底ベクトルの錐結合により張られる錐包でデータを近似するが，SINMF は k 本の基底の凸結合により張られる $k-1$ 次元単体でデータを表現する．このとき，単体によって最も誤差なくデータを表現するような

†北海道大学大学院 情報科学研究科
‡科学技術振興機構 (JST) さきがけ

頂点 (基底) の配置は原点の取り方に依らない．また，NMF と SINMF の求める基底 $\boldsymbol{F}$ は一般に異なる．SINMF の拡張として，アフィン変換に対して不変で，負値データを扱える手法 (Affine-Invariant Semi-NMF)[3] も構成できることが分かってきている．

図 1 に NMF と SINMF のある平行移動に対する基底の変化の違いの様子を示す．また，図 2 に，その平行移動に対して，どれほどの基底のずれが連続的に生じるかを示す．ここで「ずれ」の尺度として $\mathrm{gap}(\boldsymbol{F}, \tilde{\boldsymbol{F}}_j, \boldsymbol{a}_j) = \|\boldsymbol{F} + \boldsymbol{a}_j \boldsymbol{1}_{\boldsymbol{k}}^{\boldsymbol{T}} - \tilde{\boldsymbol{F}}_j\|_F^2$ で定義される関数を用いる．ここで，$\boldsymbol{F}$ は平行移動前の基底，$\boldsymbol{a}_j$ は $\boldsymbol{a}$ を始点から終点に向けて 100 分割した時の j 番目の点で表されるベクトル，$\tilde{\boldsymbol{F}}_j$ は $\boldsymbol{X} + \boldsymbol{a}_j \boldsymbol{1}_{\boldsymbol{n}}^{\boldsymbol{T}}$ を分解して得られた基底である．

図 1 はデータ行列 $\boldsymbol{X} \in \mathbb{R}^{3\times 100}$ に対して，$k = 2$ として NMF と SINMF によりで得られた 2 本の基底ベクトルを二次元表示したものである．それぞれ，(a)NMF:平行移動前 (b)NMF:平行移動後のデータと基底，(c)SINMF:平行移動前，(d)SINMF:平行移動後を表す．NMF は (b)(d) に矢印で示されている平行移動ベクトル $\boldsymbol{a}$ に対し基底とデータとの相対的な距離が大きく変化しているが，SINMF の場合平行移動に対して基底とデータとの距離の変化は非常に少なく，そのデータの特徴を一貫して保持している．また図 2 より，平行移動に対して基底のずれが増加している様子が見てとれる．NMF では，一般にずれは 0 にならないが，平行移動の大きさに対して単調に増加するとは限らない．

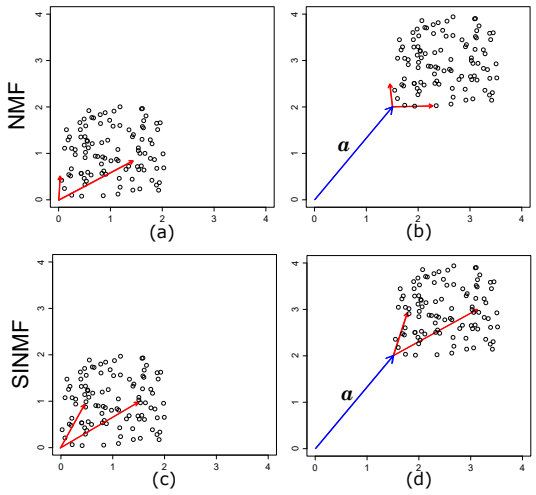

図 1: 平行移動 $\boldsymbol{a}$ の前後での基底ベクトルの変化の違い

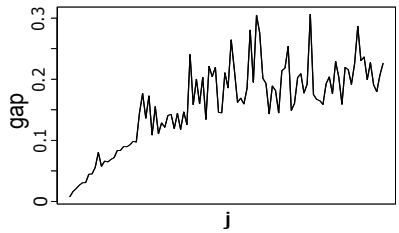

図 2: 平行移動 $j \cdot \boldsymbol{a}/100$ に対する NMF の基底のずれ

4. 数値例

実データを用いて NMF と SINMF の比較例を示す．今回は文書クラスタリングを比較の方法に採用する．データセットとして，文書クラスタリングの数値実験に汎用的に用いられているニュース文書のデータセット Reuters21578 を使用した．結果として得られる分解の解釈を容易にするため，全ての 21578 本の記事のうち，その中から原油に関する 20 本の文書，企業の合併・買収に関する 50 本の文書を選択した (この場合実質的なクラスタは 2 つとなる)．このサブセットは，文書数 70，出現した単語の種類 2010(Stop Word を除く) のデータであり，「原油関連」と「企業の合併・買収関連」の 2 つのクラスタを持つ．

このデータ行列 $\boldsymbol{X} \in \mathbb{R}^{2010\times 70}$ に対して，基底数 $k = 2$ として，$\boldsymbol{F} \in \mathbb{R}^{2010\times 2}$，$\boldsymbol{G} \in \mathbb{R}^{70\times 2}$ を求め，既知の 2 つのクラスタとの関係を見る．収束条件は，更新前後の目的関数値の誤差が定数内 $\epsilon = 10^{-7}$ かどうかで判定を行った．初めに，通常の NMF，SINMF で $\boldsymbol{X}$ の分解を行って得られた第一基底における出現頻度上位 10 語を示す．また基底はノルムが 1 になるよう正規化した．

表 1: NMF，SINMFにおけるそれぞれの $\boldsymbol{f}_1$

移動前	NMF $\boldsymbol{f}_1$
dlrs	0.432
company	0.256
mln	0.223
express	0.218
american	0.217
stock	0.212
pct	0.212
shearson	0.207
analysis	0.175
inc	0.163

移動前	SINMF $\boldsymbol{f}_1$
dlrs	0.438
pct	0.302
mln	0.274
company	0.266
inc	0.215
shares	0.214
reuter	0.209
stock	0.187
share	0.140
offer	0.140

次に，平行移動 $\boldsymbol{a} = (5, 5, ..., 5)^T \in \mathbb{R}^{2010}$（全ての単語の出現回数を 5 回増加）を加えたデータセット $\boldsymbol{X} + \boldsymbol{a}\boldsymbol{1}_{\boldsymbol{n}}^{\boldsymbol{T}}$ をそれぞれの NMF で分解した結果得られた第一基底を以下に示す．

表 2: NMF，SINMFにおける平行移動後の $\tilde{\boldsymbol{f}}_1$

移動後	NMF $\tilde{\boldsymbol{f}}_1$	$\boldsymbol{f}_1 + \boldsymbol{a}$
dlrs	0.031	0.032
pct	0.029	0.027
company	0.028	0.028
mln	0.024	0.027
shares	0.027	
inc	0.027	0.027
reuter	0.026	
stock	0.026	0.027
offer	0.025	
share	0.025	

移動後	SINMF $\tilde{\boldsymbol{f}}_1$	$\boldsymbol{f}_1 + \boldsymbol{a}$
dlrs	0.032	0.031
pct	0.029	0.029
company	0.028	0.028
mln	0.027	0.028
shares	0.027	0.027
inc	0.027	0.027
reuter	0.027	0.027
stock	0.026	0.026
offer	0.025	0.025
share	0.025	0.025

ここで，ベクトルの要素の値は特徴の中でその単語 (次元) がどれほどの重要度を占めているかを測る相対的指標にあたる．表 1，2 より通常の NMF と SINMF により生成された基底は，平行移動前，後，どちらもデータの特徴をよく表す単語が上位にとられているが,SINMF による移動後の基底は,$\boldsymbol{f}_1 + \boldsymbol{a}$ で表される真に望まれる基底に近い. ここで空白のセルは, その単語が真に望まれる基底の頻度上位 10 位に入らなかったことを意味する (順位が入れ替わっている). これにより, 通常の NMF に比べ特徴が平行移動に対し変化しない傾向が見てとれる．

5. まとめ

通常の NMF が平行移動に対して基底が変化する特性を実験を通して示し．今回新たに提案した SINMF がデータの平行移動に対して特徴を保存する点を分析した．今後の課題としては，テンソル分解への応用などが考えられる．

参考文献

[1] D. D. Lee and S. H. Sebastian, Learning the parts of objects by non-negative matrix factorization. *Nature*, vol.401, pp.788-791, 1999.

[2] D. D. Lee and S. H. Sebastian, Algorithms for Non-negative Matrix Factorization. *Advances in Neural Information Processing Systems*, 13, pp.556-562, 2001.

[3] 張 若霓, 今井英幸, Designing affine transformation based semi-nonnegative matrix factorization. 日本計算機統計学会第 30 回大会, 京都市, 2016 年 5 月 19 日-20 日.