

非負値行列因子分解の絶対値誤差最小化について

橋村 勇志¹ 山内 由紀子² 来嶋 秀治² 山下 雅史²

概要：非負値行列因子分解 (Non-negative Matrix Factorization, NMF) は、データや画像の処理にしばしば現れる。しかし、NMF の計算は NP-hard であるため、正確な解を求めることは難しい。2 乗誤差を評価尺度とする既存研究は多いが、本研究では、非負ランクの計算を動機として、絶対値誤差を評価尺度とする NMF のアルゴリズム開発に取り組む。具体的には、劣勾配法に基づく最小化アルゴリズムを提案する。提案アルゴリズムを用いて非負ランクの近似計算を行うヒューリスティックアルゴリズムを提案し、計算機実験でその性能を評価する。

キーワード：近似アルゴリズム、非負値行列因子分解

Non-negative Matrix Factorization in Absolute Error

Abstract: Non-negative Matrix Factorization (NMF) often appears in processing various datum such as image data. It is known that computing NMF is NP-hard in general, meaning that it is difficult to find an exact answer. While there are many existing works on minimization of squared error, this paper is concerned with the NMF in terms of absolute error, motivated by a computation of non-negative rank. In fact, this paper proposes a minimization algorithm based on the subgradient method. Using the algorithm, this paper proposes a heuristic algorithm to approximate non-negative ranks, and shows some results of computational experiments.

Keywords: heuristic algorithm, Non-negative Matrix Factorization

1. はじめに

非負値行列因子分解 (Non-negative Matrix Factorization, NMF) とは、 $m \times n$ 非負値行列 V が与えられたとき、 V を 2 つの非負値行列 $W(m \times a$ 行列) と $H(a \times n$ 行列) の積 WH に分解することである。例えば、 $a = \min\{m, n\}$ の場合には $V = WH$ を満たす W, H の対が存在するが、 a が小さくなると必ずしも $V = WH$ を満たす W, H の対は存在しない。 $V = WH$ が満たされる最小の a を非負ランク (non-negative rank) と呼ぶ。非負ランクの計算は NP 困難であり、従って NMF の計算も一般に NP 困難である [3]。本論文では、非負ランクの近似計算を動機として、非負値行列分解の絶対値誤差最小化を論じる。

多くの先行研究では、 V と WH を 2 乗誤差で評価している。積 $WH = X$ とした時、評価関数 $g(W, H)$ は、

$$g(W, H) = \sum_{i=1}^m \sum_{j=1}^n (v_{ij} - x_{ij})^2$$

で表される。2 乗誤差最小化においては、乗法的更新アルゴリズムについてよく研究されている。更新式は、

$$H_{an} \leftarrow H_{an} \frac{(W^T V)_{an}}{(W^T W H)_{an}},$$

$$W_{ma} \leftarrow W_{ma} \frac{(V H^T)_{ma}}{(W H H^T)_{ma}}$$

と表される。 W, H の初期値をランダムにとり、更新式を繰り返し実行すると、積 WH は V に近づく。

本研究では、絶対値誤差、すなわち

$$f(W, H) = \sum_{i=1}^m \sum_{j=1}^n |v_{ij} - x_{ij}|$$

の最小化を扱う。絶対値誤差最小化では一般に微分不可能な点が存在し、関数の微分可能性を要求するニュートン法

¹ 九州大学工学部電気情報工学科
Department of Electrical Engineering and Computer Science, Kyushu University

² 九州大学大学院システム情報科学研究院
Graduate School of Information Science and Electrical Engineering, Kyushu University

が素朴には適用できない．本研究では，劣勾配法を用いた再帰型のアルゴリズムを提案する．また，これを応用して， a の値を二分探索して行列の非負ランクを求める近似アルゴリズムを提案し，計算機実験により検証を行う．

2. 劣勾配法

定義 1. 集合 $\Omega \subseteq R^k$ 上で定義された関数 $g: \Omega \rightarrow R$ に対し，ベクトル $\mathbf{d} \in R^k$ が点 $\mathbf{x}_0 \in \Omega$ での g に対する劣勾配であるとは，任意の $\mathbf{x} \in \Omega$ に対し

$$g(\mathbf{x}_0) + \mathbf{d}^\top (\mathbf{x} - \mathbf{x}_0) \leq g(\mathbf{x})$$

が成り立つことである．また， $\mathbf{x}_0$ における g の劣勾配の集合を $\partial g(\mathbf{x}_0)$ と書き，劣微分という [2]．

集合 $\Omega \subseteq R^k$ 上で定義された関数 $g: \Omega \rightarrow R$ の最適化問題 $\min\{g(\mathbf{y}) \mid \mathbf{y} \in \Omega\}$ を考えると，劣勾配法のアルゴリズムは以下のものである．ここで， Ω 中の点で $\mathbf{y}$ からの距離が最も小さい点を $P_\Omega(\mathbf{y})$ で表す．

アルゴリズム 1.

- (1) 初期ベクトル $\mathbf{y}_1 \in \Omega$ を適当なものに固定する． $i = 1$ とする．
- (2) 劣勾配 $\mathbf{d}_i \in \partial g(\mathbf{y}_i)$ を求め， $\hat{\mathbf{y}}_i := P_\Omega(\mathbf{y}_i - s_i \mathbf{d}_i)$ を計算する．
- (3) $\mathbf{y}_i = \hat{\mathbf{y}}_i$ ならば， $\mathbf{y}_i$ を出力して終了する．
- (4) $\mathbf{y}_i \neq \hat{\mathbf{y}}_i$ ならば， $\mathbf{y}_{i+1} := \hat{\mathbf{y}}_i, i = i + 1$ と更新し，(2) に戻る．

ここで，数列 $\{s_i\}$ は， $s_i = 1/i$ がよく用いられる [2]．劣勾配法を用いて，微分不可能な連続凸関数を最小化する．

3. 絶対値誤差最小解の満たす性質

絶対値誤差を最小にする解は次の定理を満たす．

定理 1. 与えられた $m \times n$ 非負値行列 V と正整数 a に対し，非負値行列 $W(m \times a \text{ 行列})$ と $H(a \times n \text{ 行列})$ の対は， $f(W, H) = \|V - WH\|_1$ を最小にするものとする．この時， $Y := V - WH$ には少なくとも $\min\{m, n\}$ 個の 0 要素をもつ．

定理 2. 与えられた $m \times n$ 非負値行列 V と正整数 a に対し，非負値行列 $W(m \times a \text{ 行列})$ と $H(a \times n \text{ 行列})$ の対は， $Y := V - WH \geq 0$ の条件の下で $f(W, H) = \|V - WH\|_1$ を最小にするものとする．この時， Y には少なくとも $\min\{m, n\}$ 個の 0 要素が存在する．

なお，定理 1, 2 は絶対値誤差に固有の性質で，2 乗誤差には当てはまらない．

命題 1. 行列 Y を， $Y = V - WH$ で定義すると，2 乗誤差の場合，誤差が最小になる Y に 0 要素が少なくとも $\min\{m, n\}$ 個存在するとは限らない．

4. 絶対値誤差最小化のアルゴリズム

本章では，絶対値誤差 $f(W, H)$ を最小化するアルゴリズム

を提案する．アルゴリズムは，劣勾配法 [3] に基づく．

アルゴリズム 2.

- (1) W, H の初期値をランダムに与える． $n = 1$ とする．
- (2) W の劣勾配をもとに W を更新する．

$$W^{(n+1)} = W^{(n)} - s^{(n)} \frac{\partial f(W, H)}{\partial W} \quad \text{ただし, } s^{(n)} = \frac{1}{n}$$
- (3) H の劣勾配をもとに H を更新する．

$$H^{(n+1)} = H^{(n)} - s^{(n)} \frac{\partial f(W, H)}{\partial H} \quad \text{ただし, } s^{(n)} = \frac{1}{n}$$
- (4) $n = n + 1$ として，(2), (3) を繰り返す．
- (5) 定められた回数繰り返し出力する．

また，劣微分については，ここでは， $y = |x|$ の劣微分を

$$\frac{dy}{dx} = \begin{cases} -1 & (x < 0) \\ 0 & (x = 0) \\ 1 & (x > 0) \end{cases}$$

とする．これに対応して，説明のために符号関数 $\text{sgn}(x)$ を，

$$\text{sgn}(x) = \begin{cases} -1 & (x < 0) \\ 0 & (x = 0) \\ 1 & (x > 0) \end{cases}$$

と定める．

このとき，劣勾配ベクトルは各 W, H 行列の要素ごとに計算できる．行列 Y を $Y = V - X$ と定義し，

$$\begin{aligned} \frac{\partial f(W, H)}{\partial w_{ij}} &= \sum_{k=1}^n -h_{jk} \text{sgn}(y_{ik}) \\ \frac{\partial f(W, H)}{\partial h_{ij}} &= \sum_{k=1}^m -w_{kj} \text{sgn}(y_{ki}) \end{aligned}$$

これをもとに，要素ごとに更新を行う．

まず， $V = WH$ を満たす解 W, H に対して，アルゴリズム 2 は所望の解を出力することを次の定理で示す．

定理 3. $f(W, H) = 0$ のとき， $W^{(n+1)} = W^{(n)}$ かつ $H^{(n+1)} = H^{(n)}$ が成り立つ．

Proof. $f(W, H) = 0$ のとき，任意の i, j について， $\text{sgn}(y_{ij})$ は 0 である．これと，

$$\begin{aligned} \frac{\partial f(W, H)}{\partial w_{ij}} &= \sum_{k=1}^n -h_{jk} \text{sgn}(y_{ik}) \\ \frac{\partial f(W, H)}{\partial h_{ij}} &= \sum_{k=1}^m -w_{kj} \text{sgn}(y_{ki}) \end{aligned}$$

より，

$$\frac{\partial f(W, H)}{\partial w_{ij}} = 0, \frac{\partial f(W, H)}{\partial h_{ij}} = 0$$

が成り立つ．よって，

$$W^{(n+1)} = W^{(n)} - s^{(n)} \frac{\partial f(W, H)}{\partial W},$$

$$H^{(n+1)} = H^{(n)} - s^{(n)} \frac{\partial f(W, H)}{\partial H}$$

より,

$$W^{(n+1)} = W^{(n)}, H^{(n+1)} = H^{(n)}$$

が成り立つ. よって繰り返し後の値が元の値と同じになっている. $\square$

$\square$

一般に絶対値誤差最小化が難しいと考えられる命題を次に示す.

命題 2. 評価関数 $f(W, H)$ は, 凸関数ではない.

Proof.

$$V = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} W = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} H = \begin{pmatrix} h_1 & h_2 \end{pmatrix}$$

このとき, $f(w_1, w_2, h_1, h_2) = |1 - w_1 h_1| + |1 - w_1 h_2| + |0 - w_2 h_1| + |1 - w_2 h_2|$ が凸関数ではないことを示す.

次の2つの引数のとき, 凸関数ではないことが示される.

$x = (1, 0.5, 2, 2)$ とし, $y = (1, 1.5, 2, 1)$ とする.

このときの値は, $f(x) = 3, f(y) = 4.5$ となる. $\alpha = 0.5$ とすると,

$$f((1 - \alpha)x + \alpha y) = 4$$

となる一方で,

$$(1 - \alpha)f(x) + \alpha f(y) = 3.75$$

が成り立つ. 以上の議論から, $f((1 - \alpha)x + \alpha y) \leq (1 - \alpha)f(x) + \alpha f(y)$ が満たされないため, $f(W, H)$ は一般に凸関数でない. $\square$ $\square$

5. 非負ランクを求めるヒューリスティクス

本章では, $m \times n$ 行列 V が与えられたとき, V の非負ランクを求めるための近似アルゴリズムを与える. W, H の各要素の初期値は, 0 から 4 の整数からランダムで決まる. 最適解を求めるアルゴリズムの収束性が不完全なため, 同じ k に対して 3 回の NMF を試みている.

アルゴリズム 3.

- (1) 非負ランクの最大値を $\min\{m, n\}$, 最小値を 1 とする.
- (2) k を二分法で定める. 具体的には, 現時点の非負ランクの最大値と最小値の平均値の切り上げ整数を k とする.
- (3) $m \times k$ 行列 W , $k \times n$ 行列 H を用意する.
- (4) アルゴリズム 2 を実行する.
- (5) 1 回でも評価関数 $f(W, H) < 10^{-4}$ になっていれば, 非負ランクの最大値を k とする.
- (6) 評価関数 $f(W, H) < 10^{-4}$ になっていなければ, 非負ランクの最小値を $k + 1$ とする.
- (7) 手順 (2)~(6) を繰り返し, 非負ランクの最大値と最小値が等しくなったらその値を非負ランクとして出力する.

6. まとめ

本論文では, 非負値行列因子分解の絶対値誤差最小化について議論した. 評価関数を最小化するためのアルゴリズムを示し, 最適解に必要な条件と, 劣勾配法に基づく収束性について述べた. これを利用して近似的に非負ランクを求めるアルゴリズムを提案した. 今後の課題としては, 局所解にはまることが少ないアルゴリズムの提案と, 解となるための必要条件を厳しくすることがあげられる.

参考文献

- [1] D.D.Lee and H.S.Seung, Algorithms for non-negative matrix factorization, Proc. NIPS2000, 556-562.
- [2] 松井知己, 劣勾配法の話, <http://homepage2.nifty.com/TOMOMI/text/subg.pdf>, 1997 年 5 月
- [3] S. A. Vavasis, On the complexity of nonnegative matrix factorization, SIAM J. Optim., 20 (2009), 1364-1377