非負行列分解の絶対値誤差最小化について

橋村 勇志　　山内 由紀子　　来嶋 秀治　　山下 雅史
九州大学 工学部 電気情報工学科

1　はじめに

この論文では，非負値行列因子分解 (NMF) に対する絶対値誤差最小化について述べる．ここでは，行列の全ての要素が 0 以上である非負行列についての議論をする．

2　準備

非負値行列因子分解では，$m \times n$ 行列 V が与えられたとき，２つの行列 W, H を用いて $V \fallingdotseq WH$ と近似することを考える．このとき，a は与えられるパラメータとし，W は $m \times a$ 行列，H は $a \times n$ 行列とする．ただし、$a < m$，$a < n$ とする．また，便宜のため $X = WH$ と表す.

本稿では，W, H を求めるアルゴリズムを以下に提案する．

3　アルゴリズム

まず，$X = WH$ で求められる行列 X と，元の行列 V の誤差について議論する．誤差の評価関数 $f(W, H)$ を,

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

と定義する．すなわち，絶対値誤差である．

本稿では，絶対値誤差 $f(W, H)$ の値を最小化するアルゴリズムを提案する．アルゴリズムは、劣勾配法に基づく．

(1)W,H の初期値を定める．$n = 1$ 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．

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

(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

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\begin{cases}-1 & (x<0)\\ 0 & (x=0)\\ 1 & (x>0)\end{cases}$$

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

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

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，

$$\frac{\partial f(W,H)}{\partial w_{ij}}=\sum_{k=1}^{n}-h_{jk}\mathrm{sgn}(y_{ik})$$

$$\frac{\partial f(W,H)}{\partial h_{ij}}=\sum_{k=1}^{m}-w_{kj}\mathrm{sgn}(y_{ki})$$

これをもとに，要素ごとに更新を行う.

定理 1
$f(W,H)=0$ のとき，W,H の値は，前回の繰り返しでの値と同じである.

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

$$\frac{\partial f(W,H)}{\partial w_{ij}}=\sum_{k=1}^{n}-h_{jk}\mathrm{sgn}(y_{ik})$$

$$\frac{\partial f(W,H)}{\partial h_{ij}}=\sum_{k=1}^{m}-w_{kj}\mathrm{sgn}(y_{ki})$$

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．

(証明終)

4 計算機実験

次の計算機環境で実験を行った．
OS　Windows7
言語　C++
マシンスペック　CORE i5
コンピュータにプログラムを実装し，シミュレーションを行ったところ，繰り返し回数は，$n = 10$ 万回だと，最適解が得られないことが 50%ほどあった．しかし，$n = 100$ 万回だとほとんどの場合に最適解が得られた．よって，$n = 100$ 万とした．

ただ，一部の場合において，n の回数を増やしても，劣勾配ベクトルが 0 になってしまい，局所的解にはまり最適解が得られないことがあった．

5 まとめ

絶対値誤差を最小化するアルゴリズムを提案し，証明を行い，計算機実験で検証を行った．検証は概ね成功したが，一部において不十分な点が発見された．それについては，今後の課題とする．

参考文献

[1]　D.D.Lee and H.S.Seung, Algorithms for non-negative matrix factorization, Proc. NIPS2000, 556–562.