

# 奇异值分解与非负矩阵分解 在数据降维方面的特性分析

徐利民<sup>1</sup>，龚珊<sup>1</sup>，余再军<sup>2</sup>

(1. 信息工程大学信息工程学院电子科学与技术系，河南郑州 450002， 2. 郑州铁路局客车车辆段技术科，河南郑州 450004)

**摘要：**奇异值分解（SVD，Singular Value Decomposition）和非负矩阵分解（NMF，Non-negative Matrix Factorization）是两种常见的数据降维方法。分析表明两种算法在其数据表示意义、表示能力、数据重建计算量、数据存储量等方面有不同的特性。

**关键词：**信号与信息处理 奇异值分解 非负矩阵分解

## Characteristics of Singular value decomposition and non-negative matrix factorization for data reduction

XU Li-min<sup>1</sup>，GONG Shan<sup>1</sup>，YU Zai-jun<sup>2</sup>

(1.Information Engineering University, Zhengzhou 450002, China, 2. Zhengzhou Railway Bureau Bus Depot Technology Section,Zhengzhou 450004)

**Abstract:** Singular value decomposition (SVD) and non-negative matrix factorization (NMF) are two common methods of data reduction. In this paper, some aspect of the two algorithms,such as the meaning of the decomposition data, the explanation ability, computation amount of data reconstruction, data storage and so on were analyzed.

**Key words:** signal and information processing singular value decomposition non-negative matrix factorization

### 1. 引言

1999年，《nature》杂志发表了D.D.Lee和H.S.Seung的论文[1]，最早提出了用非负矩阵分解（NMF，Non-negative Matrix Factorization）方法进行大规模数据降维处理的新思想，引起了广大科研人员的密切关注。奇异值分解（SVD，Singular Value Decomposition）是一种较早出现的数据降维与压缩算法。本文将两种不同分解方法进行对比，在Frobenius范数意义下，研究比较两种分解在数据降维的数据量大小、准确性、对不同数据类型的适应能力、矩阵的稀疏性与正交性等方面的特点。

### 2. 奇异值分解

奇异值分解算法将矩阵X 分解为

$$X_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T \quad (1)$$

其中U和V是正交矩阵， $\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}$ ， $\Sigma_r = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$ ，是对角矩阵，其中

$r = \min(m, n, \text{rank}(X))$ 。其中 $\sigma_i = \sqrt{\lambda_i}$  ( $i=1, 2, \dots, r$ ) 称为矩阵X的奇异值， $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$  对应于矩阵 $AA^T$ 与 $A^TA$ 的非零特征值，矩阵U和V的列向量分别是对应于矩阵 $AA^T$ 与 $A^TA$ 的非零特征值的特征向量。

用该分解可得到矩阵的低秩逼近（截断的SVD表示）：

$$X_k = U_k \Sigma_k V_k^T \quad (2)$$

在  $X$  所有的秩为  $k$  的近似中，在Frobenius范数意义下， $X_k$  是  $X$  的全局意义上的最佳近似，即  $\|X - X_k\|_F = \min_{\text{rank}(Y)=k} \|X - Y\|_F$ 。

奇异值分解带来数据压缩能力正是基于此。记录SVD分解的较大  $k$  个奇异值和  $U$ 、 $V$  矩阵与这  $k$  个奇异值相对应的奇异向量，可以重建原数据阵在Frobenius范数意义下的有损近似。

如果把式(1)写为：

$$X_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_r u_r v_r^T \quad (3)$$

由(3)式可以看出， $X$  是  $u_1 v_1^T$ 、 $u_2 v_2^T$ 、 $\dots$ 、 $u_r v_r^T$  的线性组合， $\sigma_1, \sigma_2, \dots, \sigma_r$  是加权系数，

也可以说  $\sigma_1, \sigma_2, \dots, \sigma_r$  是矩阵  $X$  在  $u_1 v_1^T$ 、 $u_2 v_2^T$ 、 $\dots$ 、 $u_r v_r^T$  构成的基空间的投影坐标。但是不同数据矩阵会产生不同的基与坐标的组合形式，数据信息在SVD分解后将主要存在于基矩阵中，也即正交矩阵  $U$  和  $V$  具有不可分割性。实验一可以证实这一点。

实验一：选取四幅人脸图像进行SVD分解，再考察每幅图像的  $U$ 、 $V$  和  $\Sigma$  矩阵储存信息的情况，结果如图1。

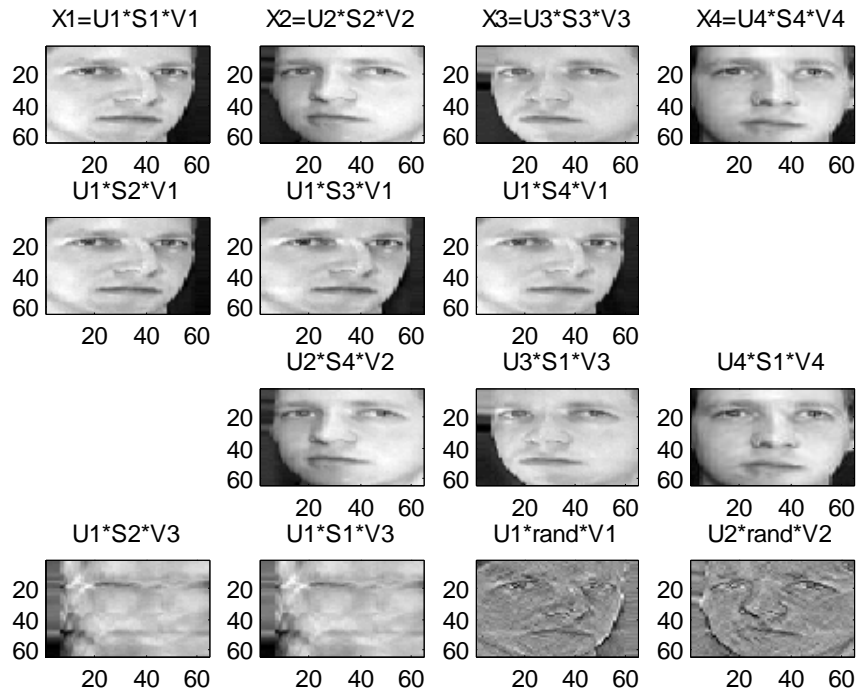


图1：第一行是四幅人脸图像  $X1$ ,  $X2$ ,  $X3$ ,  $X4$ 。分别进行奇异值分解得到  $X1=U1*S1*V1^T$ ,  $X2=U2*S2*V2^T$ ,  $X3=U3*S3*V3^T$ ,  $X4=U4*S4*V4^T$ 。第二行图像使用  $U1$ 、 $V1$  和其他图像的奇异值重建图像，发现重建的图像仍然接近  $X1$ 。第三行图像分别用  $U2$ 、 $V2$ ,  $U3$ 、 $V3$ ,  $U4$ 、 $V4$  组合其他奇异值重建图像。第四行前两幅图像是用  $U1$ 、 $V3$  和  $S2$ ,  $U1$ 、 $V3$  和  $S1$  重建的图像，后两幅分别用  $U1$ 、 $V1$  和选取的随机数奇异值重建数据  $X1$  以及  $U2$ 、 $V2$  和随机数奇异值重建数据  $X2$  的情况。

结论是： $U$ 、 $V$  矩阵携带了图像的主要信息，并且二者不可分割。

### 3. 非负矩阵分解

非负矩阵分解,也是一种数据降维方法。它的基本想法就是将一个非负矩阵近似分解为两个非负矩阵的乘积:

$$X_{m \times n} \approx U_{m \times k} V_{k \times n} \quad U \geq 0, V \geq 0 \quad (4)$$

其中 $k < m, n$ 。非负矩阵分解的算法是一个带约束的非线性规划问题(假设 $U$ 的列数 $k$ 已知):

$$\min \|X - UV\|_F \quad st. U > 0, V > 0 \quad (5)$$

其解与初值关系很大,通常只能得到局部极小值,而不是像SVD那样得到全局最优的解。

定理1给出了非负矩阵分解的基本迭代算法<sup>[2]</sup>:

定理1: 更新规则:  $V_{a\mu} \leftarrow V_{a\mu} \frac{(U^T X)_{a\mu}}{(U^T UV)_{a\mu}}; U_{ia} \leftarrow U_{ia} \frac{(XV^T)_{ia}}{(UVV^T)_{ia}}$ 可以保证范数 $\|X - UV\|$ 递减。

其中,  $1 \leq i \leq m, 1 \leq a \leq k, 1 \leq \mu \leq n$ 。

文献[3]用概率模型描述非负矩阵分解的算法: 将矩阵看成含加性噪声的线性混合体模型:

$$X_{m \times n} = U_{m \times k} V_{k \times n} + E_{n \times m} \quad (6)$$

根据噪声服从不同的概率分布,可以得到不同类型的目标函数,比如定理一所提到的算法就是将噪声看作高斯噪声,用最大似然法求解的结果。

文献[4]将非负矩阵分解算法归类为两大类: BNMF (Basic NMF基本NMF算法)和INMF (Improved NMF改进的NMF算法),其中BNMF又分为基于单个目标函数的算法和基于目标函数族的算法,INMF细分为三类:(1)稀疏性增强NMF算法(2)鉴别性NMF算法(3)加权NMF算法。

非负矩阵分解后的矩阵仅含有非负元素,通过改进的算法可以使矩阵变得稀疏,占用存储空间少,恢复数据的运算具有纯加性,运算速度快,原矩阵 $X$ 中的列向量可以解释为对 $U$ 矩阵中所有列向量(称为基向量)的加权和,权重系数为 $V$ 矩阵中对应列的元素。这种基于基向量组合的表示形式可基于局部地反映数据潜在的线性结构,具有很好的可解释性,反映了人类思维中“局部构成全局”的概念。

实验二: 100幅人脸图像,每幅 $32 \times 32$ 像素。进行NMF分解为 $U$ 、 $V$ 矩阵( $k=16$ ), $V$ 矩阵每一列是对 $U$ 矩阵基图像的加权线性组合。

1

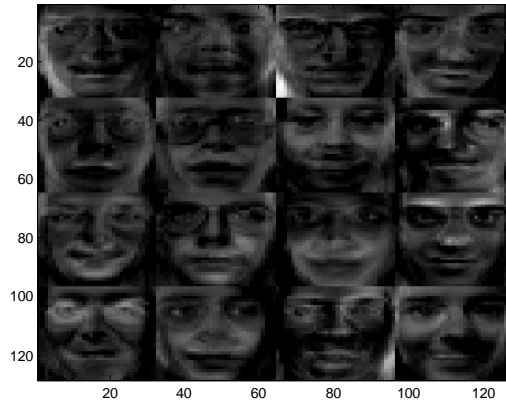


图2 基矩阵 $U$ 中图像

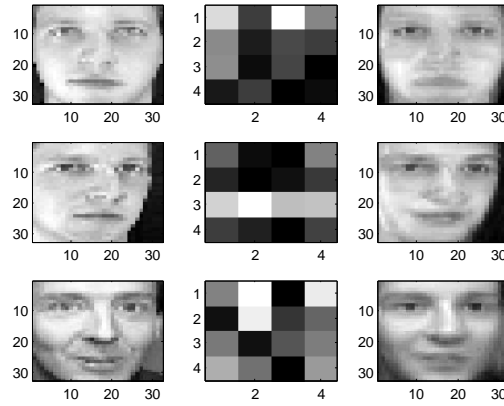


图3 还原过程（左侧为原图像，中间列是对应V中权值，右侧是还原的图像）

U矩阵是NMF分解所得基矩阵，具有稀疏性（如果在NMF算法上增加基矩阵列矢量的正交性约束会得到更具稀疏性的分解），V矩阵每一列是重建图像的权矢量。可以看到，NMF分解确实可以将图像数据的局部特征提取出来。

#### 4. SVD与NMF的比较

一般把SVD和NMF分解的U矩阵中的列称为特征向量。从形式上看，二者的特征向量区别明显。NMF的特征向量由于具有非负的特点，特征向量之间的内积必大于零，不可能完全正交，说明NMF分解的特征向量存在信息冗余。而SVD分解的特征向量彼此正交，但失去了非负的特点，可解释性变差。

SVD分解的特征向量在不同的维度都能够对样本向量做最好的近似，使近似后的结果与样本的误差在欧式距离定义下达到最小。NMF在分解过程中基数目需要预先确定，每次分解只是在一个确定维度下对样本的局部近似，其与样本的近似程度只能在欧式距离上达到局部最小而非全局最小，受初始值影响明显。

在体现样本特征方面，假设SVD所抽取的 $u_i$ 对应最大奇异值 $\sigma_i$ ，则样本在 $u_i$ 上的投影能量总和为 $E_i = (X^T u_i)^T (X^T u_i) = \sigma_i^2$ 。这样对应奇异值越大的向量具有更大的样本投影能量，在体现样本特性方面有主次之分。而NMF所抽取的特征向量彼此间重要程度差别不大，无主次之分。

NMF的非负特性使它所抽取的特征向量有很好的物理解释，比如图像领域，一个非负特征向量可以解释为一幅特征图，每个元素代表一个像素。文本领域，一个非负特征向量可以解释为一个“主题”，每个元素代表某个单词在主题中的重要程度。SVD所抽取的特征向量由于存在负值就不能有这些物理解释。

根据式(2),SVD分解需要存储的数据量为：U矩阵 $m \times k$ ，V矩阵 $k \times n$ ， $\Sigma$ 阵只存储 $k$ 个主对角奇异值，共 $(m \times k) + k + (k \times n)$ 数据。重建原数据阵按照式（3）方法计算，

$$\sum_{i=1}^k \sigma_i u_i v_i^T$$

中计算 $u_i v_i^T$ 需要 $mn$ 次乘法，之前需要 $\sigma_i$ 先和 $u_i$ 、 $v_i^T$ 的数据较短者计算

$k \times \min(m, n)$ 次乘法，所以共需要计算 $kmn$ 次加法和 $kmn + k \times \min(m, n)$ 次乘法。

而式(4)的NMF分解需要存储 $(m \times k) + (k \times n)$ 数据。当NMF分解导致U矩阵稀疏时，可以减少所记录的数据量。但考虑到稀疏矩阵记录数据需要增加地址数据，所以只有U矩

阵有明显稀疏性时（如非零元少于 $\frac{2}{3}(m \times k)$ 时）才可能缩小记录数据量。而数据还原时，如果根据下式计算：

$$X_{m \times n} \approx U_{m \times k} V_{k \times n} = \sum_{i=1}^k u_i v_i^T \quad (7)$$

其中： $u_i = [u_{i1} \ u_{i2} \dots u_{im}]^T$   $v_i = [v_{i1} \ v_{i2} \dots v_{in}]^T$

则需要 $kmn$ 次加法和 $kmn$ 次乘法。不过和SVD分解的数据重建不同，NMF分解的数据具有非负性，重建时不需使用减法。

实验三：相同存储量下二者对原数据的还原质量。利用 $\|X - X_k\|_F$  考察数据还原质量。

由于低秩逼近SVD分解的数据记录量只比NMF略多 $k$ ，所以以 $k$ 为数据记录量参数。

图4数据采用70\*60的随机数据矩阵，图5数据采用64\*64的人脸图像数据。实验四相同。

（采用了nmf toolbox中的五种典型算法，包括als、alsobs、mm、cjlin、prob）

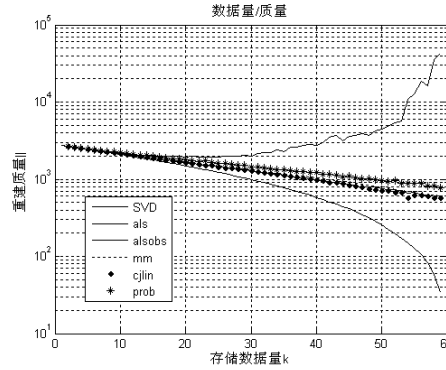


图4 对70\*60的随机数据矩阵还原质量随数据记录量的变化情况

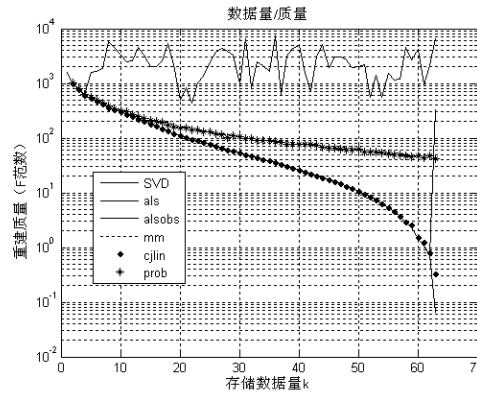


图5 对64\*64的人脸图像数据还原质量随数据记录量的变化情况

从实验结果看，各种NMF算法具有不同的重建质量，实际上采用不同的初始值计算则每次结果也不相同。NMF算法得到的局部最优解差于SVD的全局最优解。但对人脸数据的处理显示部分算法其性能接近SVD（图5所示）。

SVD分解与 $k$ 无关，便于重建数据时合理选择 $k$ 值；NMF分解须指定 $k$ 值进行，且分解过程是迭代运算，耗时较长，不便于合理选择 $k$ 值。

实验四：各种NMF算法U、V矩阵稀疏性。图6、图8数据采用70\*60的随机数据矩阵，图7、图9数据采用64\*64的人脸图像数据。

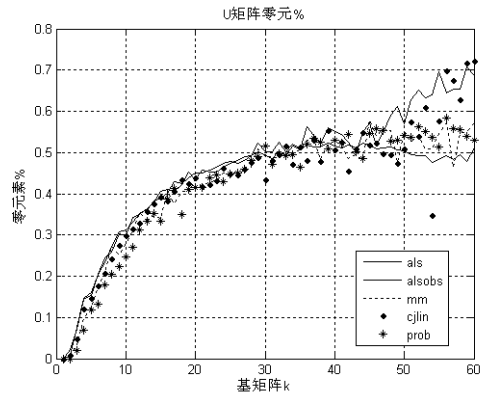


图6 70\*60的随机数据矩阵U矩阵稀疏性

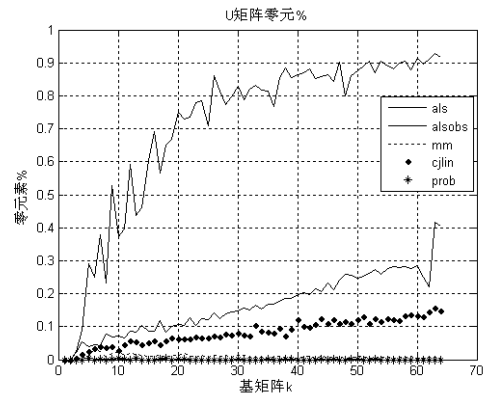


图7 64\*64的人脸图像数据矩阵U矩阵稀疏性

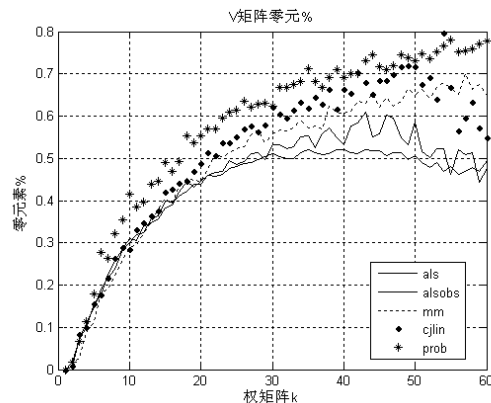


图8 70\*60的随机数据矩阵V矩阵稀疏性

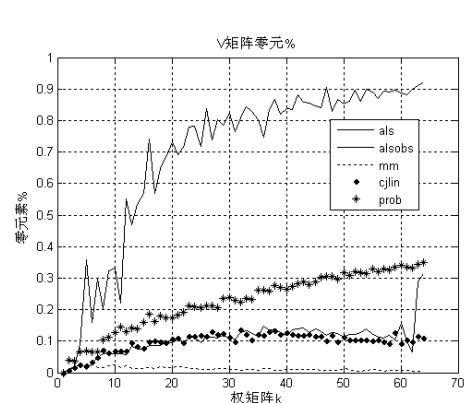


图9 64\*64的人脸图像数据矩阵V矩阵稀疏性

实验结果表明, 随机数矩阵的NMF分解零元较多, 采用稀疏矩阵记录可以明显节省数据存储量(图6、图8)。但在更具有应用特性的人脸数据中, 除als算法外其他算法零元明显较少, 并且, 各种算法的分解表明NMF分解并非具有U矩阵零元多则V矩阵零元少的相关特性(图7、图9)。

## 5. 结束语

本文将两种不同分解方法进行对比, 在Frobenius范数意义下, 研究比较两种分解在数据降维的数据量大小、准确性、对不同数据类型的适应能力、矩阵的稀疏性与正交性等方面的特点。结果表明, SVD分解具有全局最优意义上的处理能力, 但数据的可解释性不强; NMF需要迭代计算, 运算速度较慢, 但分解的可解释性强, 带有辅助约束的算法可使分解的矩阵有更大的稀疏性。另外, NMF分解针对不同类型的数据进行处理性能差异明显。

## 参考文献

- [1] Daniel D. Lee, H. Sebastian Seung, Learning the parts of objects by non-negative matrix factorization[J]. Nature 401 (6755): 788-791.
- [2] Daniel D. Lee, H. Sebastian Seung, Algorithms for Non-negative Matrix Factorization[A]. Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference. MIT Press. pp. 556-562.
- [3] 刘维湘, 郑南宁, 游屈波, 非负矩阵分解及其在模式识别中的应用[J]. 科学通报, 第51卷第3期, 2006年2月, pp241-250
- [4] 李乐章, 毓晋, 非负矩阵分解算法综述[J], 电子学报, Vol36 No.4 Apr 2008, pp737-743
- [5] 郭勇, 鲍丽春, 基于非负矩阵分解的特征向量抽取方法特点研究[J], 计算机工程与应用, 2007, 43(17), pp154-159

**作者简介：**

徐利民 (1974, 07), 男, 河南郑州人, 硕士, 讲师, 信息工程大学, 主要研究方向: 信号与信息处理。

龚珊 (1977, 09), 男, 湖北随州, 硕士, 讲师, 信息工程大学, 主要研究方向: 电路与系统

余再军 (1972, 06), 男, 河南郑州人, 学士, 高级工程师, 郑州客车车辆段, 主要研究方向: 自动控制