

# 矩阵奇异值分解在数据压缩中的应用

## Applications of Matrix Singular Value Decomposition in Data Compression

金青海 晏祖根

Qinghai Jin Zugen Yan

江西工程学院 中国·江西 新余 338000

Jiangxi University of Engineering, Xinyu, Jiangxi, 338000, China

**摘要:** 矩阵奇异值分解是在机器学习中广泛应用的算法, 它不仅可以用于数据降维算法中的特征分析, 还可用于推荐系统中的数据处理, 以及自然语言处理等领域。它是很多机器学习算法的基础。论文详述了矩阵奇异值分解的原理及发展历程, 基于该方法可构造相应的数据压缩算法, 该算法具有一定的应用和理论研究价值。

**Abstract:** Matrix singular value decomposition is an algorithm widely used in machine learning. It can not only be used for feature analysis in data dimensionality reduction algorithms, but also for data processing in recommender systems and natural language processing. It is the basis of many machine learning algorithms. This paper describes the principle and development process of matrix singular value decomposition in detail. Based on this method, some corresponding data compression algorithms can be constructed, which has certain applications and theoretical research values.

**关键词:** SVD; EVD; 数据压缩; 误差分析

**Keywords:** SVD; EVD; data compression; error analysis

**DOI:** 10.12346/csai.v1i3.7747

## 1 引言

Beltrami 在 1873 年发现了方阵的奇异值分解 (Singular Value Decomposition, SVD), 20 世纪 30 年代 Eckart 和 Young 将其推广到了长方阵, 将 SVD 发展成为一个实际的计算工具主要得益于 Gene Golub, 文献<sup>[1]</sup>展示了 SVD 技术的各种不同的应用。

SVD 的一个最直接的应用是计算矩阵乘积  $A^T A$  的特征值分解 (eigenvalue decomposition, EVD), 在主成分分析 (Principal Component Analysis, PCA) 和密码分析经常遇到这类问题。

现在大量的图像文件容量一般都比较大, 图像文件的存储和传输, 在普遍的互联网环境下会产生比较大的影响。采用数字图像的压缩技术可以从一定程度上解决这些问题。目前比较成熟的图像压缩算法比较多, 各种算法都有各自的特点。

现在进行图像压缩, 可以使用专门的图像压缩软件, 但是数据压缩效率不是很理想。实现数据压缩, 需要在保证质量的前提下进行。提高压缩效率, 需要采用合适的方法, 这是很有意义的问题。现有的种类繁多的数据压缩编码方法, 都或多或少地存在一些自身发展的瓶颈问题, 有必要开发新的数据压缩编码方法。矩阵奇异值分解技术可作为一种不错的选择, 通过对矩阵的奇异值分解, 可以把任意的矩阵降为一种较小的可逆矩阵, 然后再对可逆矩阵进行处理。利用矩阵奇异值分解, 可以实现减小矩阵的维度, 减小表示原始矩阵数据所需要的数据量, 减小原始矩阵所占用的空间, 从而达到数据约简的目的。

## 2 EVD

**定理 1** 如果  $A$  是一个实对称  $n$  阶方阵, 则存在  $n$  阶正交矩阵  $V$  和对角矩阵  $\Lambda$ , 使得  $A = V\Lambda V^T$ 。其中,  $V$  的列是  $A$  的特征向量, 它们构成  $\mathbb{R}^n$  的一组标准正交基,  $\Lambda$  的对角元

【作者简介】金青海 (1987-), 男, 中国江西九江人, 硕士, 从事科学与工程计算研究。

是  $A$  的特征值,  $V\Lambda V^T$  称为  $A$  的 EVD。

### 3 SVD

定理 2 如果  $A$  是一个  $m \times n$  实矩阵,  $m \geq n$ , 则有  $A=U\Sigma V^T$ , 其中  $U$  和  $V^T$  是正交矩阵,  $\Sigma$  是对角矩阵, 即  $U^T U=I_m, VV^T=I_n, U$  是  $m \times m$  方阵,  $V$  是  $n \times n$  方阵, 是一个  $m \times n$  对角矩阵。  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$ ,  $\sigma_i$  称为  $A$  的奇异值, 非零奇异值的个数等于  $A$  的秩, 如果  $A$  是奇异的, 那么至少  $\sigma_n=0$ 。实际上, 如果  $A$  是“接近奇异”的, 那么它的部分奇异值很小。  $U$  和  $V$  的列分别称为  $A$  的左、右奇异向量。

$$\Sigma = \begin{pmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_{n-1} & & \\ & & & & \sigma_n & \\ & & & & & 0 & \\ & & & & & & \sigma_n \end{pmatrix}$$

例 1, 如果

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \\ 17 & 18 & 19 & 20 \end{pmatrix}$$

则它的 SVD 为:

$$A=U\Sigma V^T = \begin{pmatrix} -0.0965 & -0.7686 & 0.6320 & 0.0164 & -0.0186 \\ -0.2455 & -0.4896 & -0.6212 & -0.5535 & -0.0882 \\ -0.3945 & -0.2107 & -0.3165 & 0.6286 & 0.5519 \\ -0.5435 & 0.0683 & -0.0313 & 0.3377 & -0.7649 \\ -0.6924 & 0.3472 & 0.3370 & -0.4292 & 0.3197 \end{pmatrix} \begin{pmatrix} 53.5202 & 0 & 0 & 0 \\ 0 & 2.3634 & 0 & 0 \\ 0 & 0 & 0.0000 & 0 \\ 0 & 0 & 0 & 0.0000 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} -0.4430 & -0.4799 & -0.5167 & -0.5536 \\ 0.7097 & 0.2640 & -0.1816 & -0.6273 \\ -0.0049 & -0.4017 & 0.8181 & -0.4115 \\ 0.5477 & -0.7339 & -0.1753 & 0.3615 \end{pmatrix}$$

如上式所示,  $\sigma_3=\sigma_4=0$ , 表明矩阵非满秩。

### 4 SVD 和 EVD 的关系

假定有奇异值分解  $A=U\Sigma V^T$ , 则  $A^T A=(V\Sigma^T U^T)(U\Sigma V^T)=V\Sigma^T \Sigma V^T, AA^T=(U\Sigma V^T)(V\Sigma^T U^T)=U\Sigma \Sigma^T U^T$ 。由于  $\Sigma$  是对角矩阵, 故  $\Sigma^T \Sigma$  和  $\Sigma \Sigma^T$  都是对角方阵,  $A^T A=V\Sigma^T \Sigma V^T$  是  $A^T A$  的 EVD,  $AA^T=U\Sigma \Sigma^T U^T$  是  $AA^T$  的 EVD。显然,  $A$  的任何 SVD 中, 右奇异向量 ( $V$  的列) 是  $A^T A$  的特征向量, 左奇异向量 ( $U$  的列) 是  $AA^T$  的特征向量。可以证明  $A^T A$  和  $AA^T$  有相同的特征值, 具体过程如下:

证明: 当  $A \in \mathbb{R}^{m \times n} (m \geq n)$  时, 由奇异值分解  $A=U\Sigma V^T$  得  $AV=U\Sigma V^T V=U\Sigma$ , 令  $V=(v_1, v_2, \dots, v_n)$ ,

$U=(u_1, u_2, \dots, u_m)$ , 则  $A(v_1, v_2, \dots, v_n)=(u_1, u_2, \dots, u_m)\Sigma$ , 所以:

$$Av_i = \sigma_i u_i, \quad i=1, 2, \dots, n \quad (1)$$

因为  $A^T = V\Sigma^T U^T$ , 所以  $A^T U = V\Sigma^T U^T U = V\Sigma^T$ ,  $A^T(u_1, u_2, \dots, u_m)=(v_1, v_2, \dots, v_n)\Sigma^T$ , 故:

$$A^T u_i = \sigma_i v_i, \quad i=1, 2, \dots, n \quad (2)$$

由,  $A^T Av_i = A^T \sigma_i u_i = \sigma_i (A^T u_i)$ , 再由,  $A^T Av_i = \sigma_i (\sigma_i v_i) = \sigma_i^2 v_i$ 。

同理  $AA^T u_i = \sigma_i^2 u_i$ , 这说明对称矩阵  $A^T A$  和  $AA^T$  的特征向量  $v_i$  和  $u_i$  分别对应着相同的特征值  $\sigma_i^2$ ,  $A$  的奇异值是对称矩阵  $A^T A$  和  $AA^T$  的非零特征值的平方根。

特别地, 如果  $A$  本身是对称矩阵, 那么  $A$  每一个对应于特征值  $\lambda$  的特征向量也是  $A^2 = A^T A = AA^T$  的特征值  $\lambda^2$  的特征向量,  $A$  的左、右奇异向量和它的特征向量相同,  $A$  的奇异值是其特征值的绝对值, 此时  $A$  的 EVD 和 SVD 本质上是相同的, 当  $A$  没有负特征值时完全相同。对任意  $A, A^T A$  的 SVD 和 EVD 相同。

由 SVD 和 EVD 的关系知道理论上可通过形成  $A^T A$ , 计算  $A^T A$  的特征值和特征向量等一系列步骤得到  $A$  的 SVD。实际上,  $A$  中任何元素的误差在计算  $A^T A$  时被“放大”, 造成大量的精度损失, 应该避免  $A^T A$  的计算<sup>[2]</sup>。现在已有快速的迭代算法, 可计算精确度很高的奇异值和奇异向量。SVD 的许多应用往往是较准确地计算  $A^T A$  的 EVD, 具体步骤如下:

算法 1 实矩阵 EVD 的计算方法

输入:  $A \in \mathbb{R}^{m \times n}$

输出:  $A^T A$  的 EVD

算法: 判断  $A$  是否是对称矩阵:

如果是, 则直接对  $A$  作特征值分解  $A=V\Lambda V^T$ , 返回  $A^T A (A^2)$  的特征值为对角矩阵  $\Lambda$  对角元的平方, 特征向量为  $V$  的列, 算法结束;

否则, 对  $A$  作奇异值分解  $A=U\Sigma V^T$ , 返回  $A^T A$  的特征值为  $n$  阶方阵  $\Sigma^T \Sigma$  的对角元, 特征向量为  $V$  的列<sup>[3]</sup>。

### 5 关于奇异值分解的两个性质

若  $A \in \mathbb{R}^{m \times n}$ , 则有:

- ①  $\|A\|_2 = \sigma_1$ ;
- ②  $\|A\|_F = \sqrt{\sigma_1^2 + \dots + \sigma_p^2}$ 。

其中  $p = \min\{m, n\}$ 。

证明: ①由矩阵 2 范数定义,  $\|A\|_2 = \sqrt{\lambda_{\max}}$ ,  $\lambda_{\max}$  为  $A^T A$  的最大特征值。因为  $A^T A$  的特征值为  $\sigma_1^2 \geq \sigma_2^2 \geq \dots \geq \sigma_p^2$ ,  $\sigma_i (i=1, 2, \dots, n)$  为  $A$  的奇异值, 所以  $\|A\|_2 = \sqrt{\sigma_1^2} = \sigma_1$ 。

②由奇异值分解  $A=U\Sigma V^T$  得  $U^T AV=U^T U\Sigma V^T V$ , 根据  $U, V$  的正交性,  $U^T AV=\Sigma$ , 两边取 F 范数,  $\|U^T AV\|_F = \|\Sigma\|_F = \sqrt{\sigma_1^2 + \dots + \sigma_p^2}$ , 由 F 范数的正交不变性可知  $\|U^T AV\|_F = \|A\|_F = \sqrt{\sigma_1^2 + \dots + \sigma_p^2}$ 。

### 6 将 SVD 用于数据压缩

设  $A \in \mathbb{R}^{m \times n}$ , 若有  $A$  的奇异值分解  $A=U\Sigma V^T$ , 令  $u_i$  和  $v_i$  分别为  $U$  和  $V$  的第  $i$  列, 则  $A$  的奇异值分解可以表成正常数  $\sigma_i$  和单位向量  $u_i$ 、 $v_i$  的外积  $u_i v_i^T$  的乘积和的形式<sup>[4]</sup>, 即:

$$A = \sum_{i=1}^k \sigma_i u_i v_i^T \quad (3)$$

式可用图 1 形象地表示。

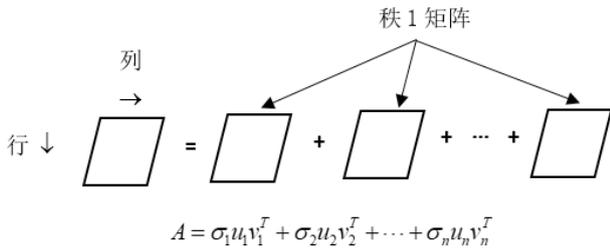


图 1 A 的奇异值分解

令  $A_i = \sigma_i u_i v_i^T$ , 则  $A = \sum_{i=1}^n A_i$ ,  $A_i$  是秩 1 的  $m \times n$  矩阵, 每个  $A_i$  仅用  $m+n$  个位置来储存, 将原矩阵通过秩 1 的外积矩阵  $u_i v_i^T$  的形式表示出来可以节省存储空间<sup>[5]</sup>。对于例 1 得到:

$$\begin{aligned}
 A &= \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \\ 17 & 18 & 19 & 20 \end{pmatrix} = U\Sigma V^T \\
 &= 53.5202 \begin{pmatrix} -0.0965 \\ -0.2455 \\ -0.3945 \\ -0.5435 \\ -0.6924 \end{pmatrix} \begin{pmatrix} -0.4430 & -0.4799 & -0.5167 & -0.5536 \end{pmatrix} \\
 &+ 2.3634 \begin{pmatrix} -0.7686 \\ -0.4896 \\ -0.2107 \\ 0.0683 \\ 0.3472 \end{pmatrix} \begin{pmatrix} 0.7097 & 0.2640 & -0.1816 & -0.6273 \end{pmatrix} \\
 &+ 0.0000 \begin{pmatrix} 0.6320 \\ -0.6212 \\ -0.3165 \\ -0.0313 \\ 0.3370 \end{pmatrix} \begin{pmatrix} -0.0049 & -0.4017 & 0.8181 & -0.4115 \end{pmatrix} \\
 &+ 0.0000 \begin{pmatrix} 0.0164 \\ -0.5535 \\ 0.6286 \\ 0.3377 \\ -0.4292 \end{pmatrix} \begin{pmatrix} 0.5477 & -0.7339 & -0.1753 & 0.3615 \end{pmatrix} \quad (4)
 \end{aligned}$$

为了压缩数据, 将较小的奇异值设置为零。如果仅使用

了 10 个奇异值, 那么这个近似  $A = \sum_{i=1}^n \sigma_i u_i v_i^T \approx \sum_{i=1}^{10} \sigma_i u_i v_i^T$ 。可以把这个思想应用于图像压缩<sup>[6]</sup>, 假设原始图像大小为  $500 \times 500$ , 需要 250 000 个数字来存储, 将所有较小的的奇异值设置为零, 仅使用前 10 ~ 20 个最大的奇异值便可得到清晰度较高的图像。近似图像通过  $U$  和  $V$  的前 10 ~ 20 列以及对应的奇异值来重建, 忽略剩下的项, 因此图像仅需 10 000-20 000 个数字存储<sup>[7]</sup>。

### 7 SVD 数据压缩算法的误差分析

若矩阵  $A$  的 SVD 为  $A = \sum_{i=1}^k \sigma_i u_i v_i^T$ , 现定义一个秩  $r$  矩阵

$$S_r = \sum_{i=1}^r \sigma_i u_i v_i^T, \text{ 用 } S_r \text{ 去近似 } A \text{ 将产生误差 } E_r = A - S_r = \sum_{i=r+1}^k \sigma_i u_i v_i^T$$

我们将 SVD 外积和形式表示的秩  $r$  矩阵  $S_r$  产生的绝对误差  $E_{ab}$  定义为:

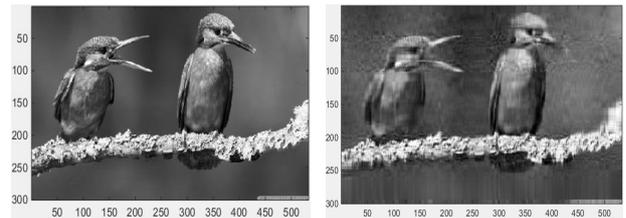
$$E_{ab} = \|E_r\|_F = \sqrt{\sum_{i=r+1}^k \sigma_i^2} \quad (5)$$

相对误差  $E_{re}$  定义为:

$$E_{re} = \frac{\|E_r\|_F}{\|A\|_F} = \left( \frac{\sum_{i=r+1}^k \sigma_i^2}{\sum_{i=1}^k \sigma_i^2} \right)^{1/2} \quad (6)$$

### 8 数据仿真实验结果及分析

论文的实验在 MATLAB 软件环境下进行, 我们选取一张分辨率为  $300 \times 533 \times 3$  的 RGB 图片, 经过图像类型变换, 转换成如图 2(a) 所示的灰度图像。



(a) 原始图像 (b) 矩阵分解后重构的图像 ( $k=30$  时)

图 2 矩阵奇异值分解的压缩效果示意图

将此灰度图作为原始图像。图 (b) 是经过矩阵奇异值分解处理后重构的图像, 观察可以发现, 图像的轮廓已经比较清晰了, 这里我们从原始图像矩阵分解后的总共 300 个奇异值中取其 1/10, 即前 30 个较大的奇异值, 而且随着  $k$  的增大, 图像的清晰度 (保真度) 会越来越高, 同时我们根据压缩比 (压缩率) 计算公式:

$$C_r = \frac{n_1}{n_2} \quad (7)$$

其中,  $C_r$  表示压缩比,  $n_1$  表示原始图像每像素的比特数,  $n_2$  表示压缩后平均每像素的比特数。按 (7) 式计算得到  $C_r = 6.3909$ , 如果对图像质量要求不是很高采用基于矩阵分解的压缩方法是可行的。

### 9 结语

矩阵奇异值分解技术有较为坚实的数学基础, 是一个强有力的数学工具, 能够达到数据压缩的效果。该算法因原理

较易理解, 便于机器实现, 且具有一定的精度, 不失为进行数据压缩的一种有效方法。

下一步的研究工作可以考虑研究矩阵奇异值分解算法的运算速度, 研究如何构造更快速的矩阵奇异值分解算法。该算法仍有进一步改进和优化的空间, 可以考虑在保证一定精度和压缩比的前提下, 进一步降低算法的时间和空间复杂度。

还可以考虑将矩阵奇异值分解算法和常用的二维快速 Fourier 变换 (FFT) 编码、二维离散余弦变换 (DCT) 编码和小波分析变换编码等图像压缩算法进行比较, 研究它们数据压缩原理的区别和联系, 研究在相同的压缩比的条件下, 比较它们的数字图像压缩结果; 可以研究在取得同样的数据压缩效果的条件下, 比较它们的压缩比的大小; 可以将矩阵奇异值分解算法的误差分析作为一个专题来研究, 这是很有意义的数值分析问题。

### 参考文献

- [1] Golub L M, Lee H M, Lehrer G, et al. Minocycline reduces gingival collagenolytic activity during diabetes: preliminary observations and a proposed new mechanism of action[J]. *Journal of periodontal research*,1983,18(5):516-526.
- [2] Chan T F, Golub G H, LeVeque R J. Algorithms for computing the sample variance: Analysis and recommendations[J]. *The American Statistician*,1983,37(3):242-247.
- [3] Golub S S. Oil prices and exchange rates[J]. *The Economic Journal*,1983,93(371):576-593.
- [4] Andrews H C, Patterson C L. Outer product expansions and their uses in digital image processing[J]. *The American Mathematical Monthly*,1975,82(1):1-13.
- [5] Long C. Visualization of matrix singular value decomposition[J]. *Mathematics Magazine*,1983,56(3):161-167.
- [6] Gabriel K R. The biplot graphic display of matrices with application to principal component analysis[J]. *Biometrika*,1971,58(3):453-467.
- [7] Moler C, Morrison D. Singular value analysis of cryptograms[J]. *The american mathematical monthly*,1983,90(2):78-87.