CUR 分解

SVD分解有一个问题，在大规模数据应用中，矩阵M大概率是稀疏的，即大多数元素是0，例如，一个表示许多文档的矩阵，其行为不同文档，列为包含的不同的单词个数，该矩阵很容易是稀疏的矩阵，因为大多数单词在大多数文档中都不会出现。同样的，一个矩阵，它的行表示顾客，列表示商品也是稀疏的因为大多数顾客不会买大多数商品。

我们不能处理高密度的矩阵，这些矩阵有百万或者数十亿的行或者列。然而使用SVD，即使M是稀疏的，U和V是密集的，Σ是对角矩阵，是稀疏的，但是Σ通常是远小于U和V的，因此它的稀疏性不会起到优化。

因此，需要提出另一种方法来压缩，叫做CUR，CUR的优点是如果矩阵M是稀疏的，那么两个大的矩阵C和R对应rows和columns，和SVD中的U和V类似，也是稀疏的。只有中间的矩阵类似于SVD中的Σ是密集的，但是这个矩阵是较小的，因此密度的影响不大。

不同于SVD，SVD可以在参数r不小于矩阵M的秩的条件下可以给出一个精确的压缩，CUR只会给出一个无论r取何值情况下的近似。已经有理论证明随着r增加，可以得到对M的收敛，但是需要将r设置尽可能大，例如1%，但是会是的方法变得不切实际。总之，通过一个相对较小的值r得到的压缩也有较好的概率是一个可用的并且精确的压缩。

CUR定义

设定矩阵M的大小为$m \times n$，选择一个目标概念值r。CUR压缩是在矩阵M中随机选择矩阵M的r列，产生一个矩阵C，大小为$m \times r$，并在矩阵M中随机选择r行，形成矩阵R，大小为$r \times n$，同样有一个矩阵U，大小为$r \times r$，矩阵U由C和R构成：

1. 设定矩阵W大小为$r \times r$，该矩阵是矩阵C和矩阵R的交集，即矩阵W的第i行和第j列的数据是矩阵M中在C中对应的第j列和在R中对应的第i行对应的数据。
2. 计算矩阵W的SVD，即$W = X\Sigma Y^T$
3. 计算Σ^+，该矩阵是对角矩阵Σ的伪逆，即如果矩阵Σ的第i个元素是$\sigma \neq 0$，那么替换为$1/\sigma$，如果为0，则不变。
4. 设定$U = Y(\Sigma^+)^2 X^T$

恰当地选择行和列

行和列的选择是随机的，然而选择影响结果的偏差，因此更重要的行和列需要有更大的机会被选择。因此我们选择$Frobenius_{范数}$来确定重要性，即求出每行或者每列元素的平方和。定义$f = \sum_{i,j} m_{ij}^2$表示矩阵M的$Frobenius_{范数}$。那么当我们选择一行时，该行对应的概率$p_i = \sum_j m_{ij}^2/f$，当我们选择一列时，该列对应的概率$q_j = \sum_i m_{ij}^2/f$。

例：

	Matrix	Alien	Star Wars	Casablanca	Titanic
Joe	1	1	1	0	0
Jim	3	3	3	0	0
John	4	4	4	0	0
Jack	5	5	5	0	0
Jill	0	0	0	4	4
Jenny	0	0	0	5	5
Jane	0	0	0	2	2

图中矩阵M的元素平方和为243，前三列的$Frobenius$范数都是51，因此每列的概率都是$51/243=0.210$。剩下的两列都是45，对应的概率为$45/243=0.185$。一共7行的$Frobenius$范数为3，27，48，75，32，50，8，因此对应的概率为0.012,0.111,0.198,0.309,0.132,0.206,0.333。

现在选择矩阵C的r列，每一列都是从矩阵M中随机选择的，但不是按照均匀分布的，每一列的选择按照概率q_j。矩阵C的每列都是从矩阵M中独立选择的，因此有些列会被多次选择，因此如何处理这种情况需要特别分析。

在选择了矩阵M的列后，将该列的元素除以$\sqrt{rq_j}$，表示除以该元素可能被选择的次数的平方根。缩放的结果作为C的列。矩阵R的组成同理。

例：

设定$r=2$，从上图的矩阵中选择，两列分别为alien和casablanca。将$[1,3,4,5,0,0,0]^T$缩放后得到$[1.54,4.63,6.17,7.72,0,0,0]^T$，这是$C$的第一列，同理第二列保留两位小数为$[0,0,0,0,9.30,11.63,4.65]^T$。

然后选择矩阵R的行，选择Jenny和Jack，因此得到的结果为$\begin{pmatrix} 0 & 0 & 0 & 5 & 5 \\ 5 & 5 & 5 & 0 & 0 \end{pmatrix}$，放缩后得到$\begin{pmatrix} 0 & 0 & 0 & 11.01 & 11.01 \\ 8.99 & 8.99 & 8.99 & 0 & 0 \end{pmatrix}$。

构造中间矩阵

最后构造矩阵U，首先使用矩阵W，矩阵W是C和R中对应的行和列在矩阵M中重叠的部分。

例：

上例中矩阵W为$\begin{pmatrix} 0 & 5 \\ 5 & 0 \end{pmatrix}$，注意首先选择的是Jenny，第一列为alien，因此W的第一行第一个元素是0，然后是5，后续同理。

矩阵U是矩阵W的伪逆，首先计算矩阵W的SVD，得到$W=X\Sigma Y^T$，然后计算Σ的伪逆Σ^+，最后$U=Y(\Sigma^+)^2X^T$。

例：

矩阵W的SVD为，

$$W=\begin{bmatrix} 0 & 5 \\ 5 & 0 \end{bmatrix}=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

求得伪逆为，

$$\Sigma^{+} = \begin{pmatrix} 1/5 & 0 \\ 0 & 1/5 \end{pmatrix} \tag{1}$$

因此U为，

$$U = Y(\Sigma^{+})^2 X^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/5 & 0 \\ 0 & 1/5 \end{bmatrix}^2 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1/25 \\ 1/25 & 0 \end{bmatrix}$$

完整的CUR分解

现在已经可以计算出矩阵C, U, R，三个矩阵的乘积是原始矩阵的近似，这个近似只有在选择了大量的行和列之后才能保证。然而直觉上讲，通过选择更重要的行和列，即使选择了较少的行和列的条件下，我们也能提炼出原始矩阵最重要的部分。

例：

$$CUR = \begin{bmatrix} 1.54 & 0 \\ 4.63 & 0 \\ 6.17 & 0 \\ 7.72 & 0 \\ 0 & 9.30 \\ 0 & 11.63 \\ 0 & 4.65 \end{bmatrix} \begin{bmatrix} 0 & 1/25 \\ 1/25 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 11.01 & 11.01 \\ 8.99 & 8.99 & 8.99 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.55 & 0.55 & 0.55 & 0 & 0 \\ 1.67 & 1.67 & 1.67 & 0 & 0 \\ 2.22 & 2.22 & 2.22 & 0 & 0 \\ 2.78 & 2.78 & 2.78 & 0 & 0 \\ 0 & 0 & 0 & 4.10 & 4.10 \\ 0 & 0 & 0 & 5.12 & 5.12 \\ 0 & 0 & 0 & 2.05 & 2.05 \end{bmatrix}$$

消除重复的行和列

多次选择某一行或者某一列是很有可能的，使用同一行两次是没有大的影响的，尽管压缩后的矩阵的秩可能小于选择的行和列的数量。反之，将其中选择的来自矩阵M的相同行或者列的元素行（设有k个重复）合并为同一行或列也是有可能的，因此使得矩阵R有更少的行，矩阵C有更少的列。然而，每一行或者每一列中剩下的向量都应该乘以$\sqrt{k}$。

在合并行或者列的时候，矩阵R的行可能比矩阵C的列更少，反之亦然。因此W不必要是一个方阵。但是我们仍然可以通过W的SVD分解求出它的伪逆，此时$W = X\Sigma Y^T$，其中的Σ是对角矩阵，但是其中某些全零行或者全零列。为了计算这样一个对角矩阵的伪逆，需要将非零元素取逆，0仍然为0，然后将矩阵转置。

例：

假定Σ为，$\Sigma = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{pmatrix}$，则求得的伪逆为$\Sigma^{+} = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1/3 \\ 0 & 0 & 0 \end{pmatrix}$。

附录

伪逆的作用

通常情况下，假定矩阵$M = XZY$，如果所有矩阵的逆都存在，则$M^{-1} = Y^{-1}Z^{-1}X^{-1}$。假设$XZY$套用到SVD的三个矩阵中，那么$X$是列正交的矩阵，$Y$是行正交的矩阵。此时，两个矩阵的逆和转置是相同的，即XX^T是一个适当大小的单位矩阵，同理YY^T也成立。因此$M^{-1} = Y^TZ^{-1}X^T$。

同时，Z是一个对角矩阵，如果对角线上没有0，那么Z^{-1}通过将矩阵Z的非0数字取逆得到。当对角线有0时，不能够直接取逆，因为无法和矩阵Z相乘得到单位矩阵。因此这就是伪逆的产生。事实上ZZ^+不是单位矩阵，而是一个对角矩阵，这个对角矩阵的对应元素在Z中对应的位置不是0，则该位置为1，对应元素为0，则该位置为0。

矩阵的逆

矩阵的逆包括两侧逆、左逆和右逆

1. 两侧逆表示满秩方阵的逆，$AA^{-1} = I = A^{-1}A$。
2. 左逆

   如果A是一个$m \times n$的列满秩矩阵，意味着A的各列线性无关，A的秩和列数相等，$r = n$，但是行可能多$m \geq n$，此时A的零空间只有0向量，并且$Ax = b$有唯一解($m = n$时)，或者无解($m > n$时)。

   对于列满秩矩阵来说，对称矩阵A^TA是一个$n \times n$的满秩方阵，因此A^TA可逆，此时

   $$(A^TA)^{-1}A^TA = I \tag{2}$$

   $$A^{-1}_{left} = (A^TA)^{-1}A^T \tag{3}$$

3. 右逆

   如果A是一个$m \times n$的行满秩矩阵，意味着A的各行线性无关，A的秩和行数相等，$r = m$，但A的列可能较多，$m \leq n$，A的左零空间只有零向量，A的零空间是$n - r$维，因此有$n - r$个自由变量，当$n > m$时，$Ax = b$有无数解，当$n = m$时，没有解。

   对于行满秩矩阵来说，对称矩阵AA^T是一个$m \times m$的满秩方阵，因此AA^T可逆，此时

   $$AA^T(AA^T)^{-1} = I \tag{4}$$

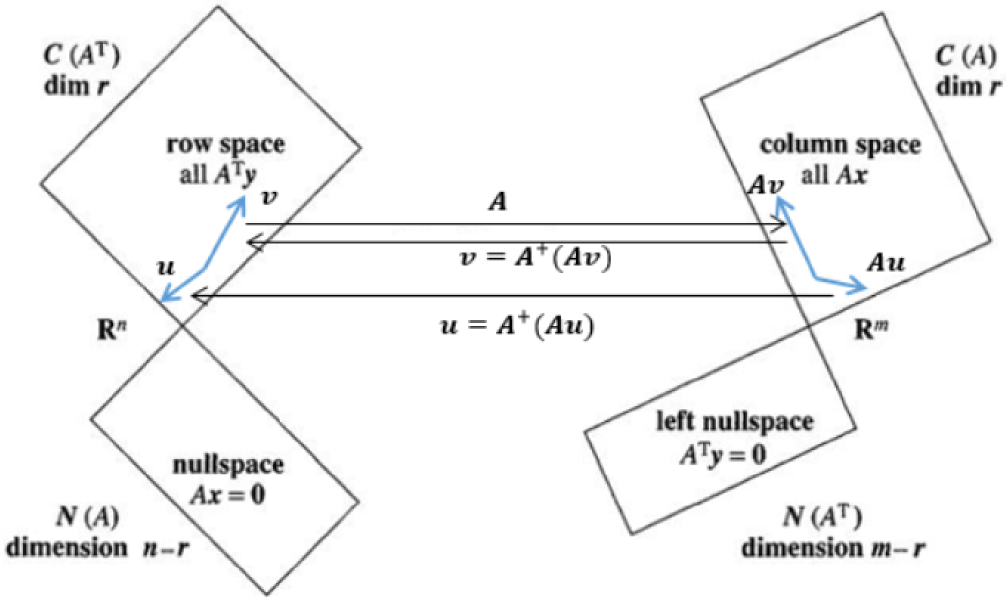
   $$A^{-1}_{right} = A^T(AA^T)^{-1} \tag{5}$$

   通常下，右乘左逆得不到单位矩阵，仅在$m = n$时才有$AA^{-1}_{left} = I$，对于列满秩的$m \times n$矩阵来说，$AA^{-1}_{left} = A(A^TA)^{-1}A^T = P$，$P$是$A$的列空间的投影矩阵，同理，左乘右逆也得不到单位矩阵，$A^{-1}_{right}A$是A的行空间的投影矩阵。

4. 伪逆

   列满秩矩阵的零空间只有零向量，有左逆矩阵，行满秩矩阵的左零空间只有零向量，有右逆矩阵；但是对于不满秩的矩阵$A(m \times n)$来说($r < m, r < n$)，两个零空间都存在，此时无法得到左逆或右逆。

   假设矩阵$A(m \times n)$是不满秩的矩阵，其行空间和列空间的维数相等，如果此时行空间的一个向量x经过A的变换，变为列空间的向量Ax，并且x和Ax是一一对应的，那么在把逆操作限制在行空间和列空间上时，A是可以进行逆操作的，A在这两个空间上的逆矩阵称为伪逆，记作A^+：



一一对应的原因： u和v是行空间的两个不同的向量，经过A的转换将变成列空间的另外两个向量Au和Av。我们假设$Au = Av$，这相当于$Au - Av = 0$，即$A(u - v) = 0$，这意味着$u - v$属于零空间。但u和v是行空间的两个向量，它们的线性组合也属于行空间，与结论矛盾，因此假设不成立，$Au \neq Av$。行空间和列空间的向量是一一对应的。

找出伪逆： $A_{m\times n}$是一个不满秩矩阵，行数和列数都大于秩，$m > r, n > r$，找出A^+的方法是利用奇异值分解，A的奇异值分解是$A = U\Sigma V^T$，则Σ^+将Σ对应不为0的元素取逆即可。

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}_{m\times n}$$

Σ和A的尺寸一致，秩为r，显然不可逆，$\Sigma^T\Sigma$和$\Sigma\Sigma^T$都不可逆，不存在左逆和右逆，只有伪逆Σ^+，

$$\boldsymbol{\Sigma}^+ = \begin{bmatrix} \frac{1}{\sigma_1} & & & & & & \\ & \frac{1}{\sigma_2} & & & & & \\ & & \ddots & & & & \\ & & & \frac{1}{\sigma_r} & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}_{n\times m}$$

伪逆是$n \times m$的矩阵，秩为r，$\Sigma^+\Sigma$和$\Sigma\Sigma^+$的结果如下

$$\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{+} = \begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}_{m\times m}, \quad \boldsymbol{\Sigma}^{+}\boldsymbol{\Sigma} = \begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}_{n\times n}$$

对于矩阵A来讲，

$$A^{+} = V\Sigma^{+}U^{T} \tag{6}$$

$$AA^{+} = U\Sigma V^{T}V\Sigma^{+}U^{T} = U\Sigma\Sigma^{+}U^{T} \tag{7}$$

结果是A的行空间的投影矩阵。

伪逆的性质

$$\begin{aligned} &AA^{+}A = A \\ &A^{+}AA^{+} = A^{+} \\ &AA^{+} = (AA^{+})^{T} \\ &A^{+}A = (A^{+}A)^{T} \end{aligned} \tag{8}$$