$\mathcal{C}$U$\mathcal{R}$ Decomposition

Anil Maheshwari

anil@scs.carleton.ca
School of Computer Science
Carleton University
Canada

Introduction

Singular Value Decomposition

- Let A be a $m \times n$ matrix of real numbers of rank r

- $A_{m\times n} = U_{m\times r}\Sigma_{r\times r}V^T_{r\times n}$, where

U is a orthonormal $m \times r$ matrix
V is a orthonormal $n \times r$ matrix
Σ is an $r \times r$ diagonal matrix and its (i,i)-th entry is σ_i for $i = 1, \ldots, r$

- Note that $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_r > 0$ and $\sigma_i = \sqrt{\lambda_i}$ where λ_i are the eigenvalues of $A^T A$

- The set of orthonormal vectors $v_1, \ldots, v_r$ and $u_1, \ldots, u_r$ are eigenvectors of $A^T A$ and AA^T, respectively. The vectors v's and u's satisfy the equation $Av_i = \sigma_i u_i$, for $i = 1, \ldots, r$

- Alternatively, we can express $A = \Sigma_{i=1}^{r} \sigma_i u_i v_i^T$

Interpreting $U, \Sigma,$ and V

Utility Matrix M as SVD $M = U\Sigma V^T$

$$M = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix} =$$

$$\begin{bmatrix} .13 & -.02 & .01 \\ .41 & -.07 & .03 \\ .55 & -.1 & .04 \\ .68 & -.11 & .05 \\ .15 & .59 & -.65 \\ .07 & .73 & .67 \\ .07 & .29 & -.32 \end{bmatrix} \begin{bmatrix} 12.5 & 0 & 0 \\ 0 & 9.5 & 0 \\ 0 & 0 & 1.35 \end{bmatrix} \begin{bmatrix} .56 & .59 & .56 & .09 & .09 \\ -.12 & .02 & -.12 & .69 & .69 \\ .40 & -.8 & .40 & .09 & .09 \end{bmatrix}$$

Issues:

1. Utility matrix M is sparse, but U and V are dense
2. Total size $= r(n+m) + r^2$
3. Interpretation of entries in U and V

A Possiblity

Can we express $M \approx \mathcal{C}U\mathcal{R}$, where

1. $\mathcal{C}$ consists of some columns of M
2. $\mathcal{R}$ consists of some rows of M
3. U is not that big
4. Square of Frobenius Norm $= \sum_{ij}(M_{ij} - (\mathcal{C}U\mathcal{R})_{ij})^2$ is small

$\mathcal{C}$U$\mathcal{R}$ Method

Let M be $m \times n$ and let $\Delta = \sum_{ij} M[i,j]^2 = ||M||_F^2$

1. For each column j, compute $p_j = \frac{1}{\Delta} \sum_{i=1}^{m} M[i,j]^2$
2. Pick α columns of M based on their probabilities (with replacement). Let $\mathcal{C}$ be the multi-set of picked columns.
3. For each element of selected columns $j \in \mathcal{C}$, scale its value to $\frac{M[*,j]}{\sqrt{\alpha p_j}}$
4. Repeat above steps for all the rows and let $\mathcal{R}$ be the multi-set of α picked and scaled rows.
5. Let W be the $\alpha \times \alpha$ matrix whose entries are from M that are common to $\mathcal{C}$ and $\mathcal{R}$
6. Construct SVD of $W = X \Sigma Y^T$
7. Construct Σ^+, where each non-zero element x of Σ is replaced by $1/x$
8. Compute $U = Y(\Sigma^+)^2 X^T$
9. Report $\mathcal{C}U\mathcal{R}$ as approximation of M

An Example

$$M = \begin{bmatrix} 4 & 1 & 1 & 0 \\ 4 & 0 & 0 & 1 \\ 0 & 0 & 5 & 5 \\ 0 & 1 & 5 & 5 \\ 0 & 1 & 5 & 3 \end{bmatrix} \qquad \sum_{ij} {M_{ij}}^2 = 171$$

Column	C_1	C_2	C_3	C_4	
$\sum_i M_{ij}$	32	3	76	60	
Row	R_1	R_2	R_3	R_4	R_5
$\sum_j M_{ij}$	18	17	50	51	35

For Rank 1 Approximation: Select C_3 and R_4 and scale them to obtain:

$$C_3 = \frac{1}{\sqrt{76/171}} \begin{bmatrix} 1 & 0 & 5 & 5 & 5 \end{bmatrix}^T = \begin{bmatrix} 1.5 & 0 & 7.6 & 7.6 & 7.6 \end{bmatrix}^T$$

$$R_4 = \frac{1}{\sqrt{51/171}} \begin{bmatrix} 0 & 1 & 5 & 5 \end{bmatrix} = \begin{bmatrix} 0 & 1.85 & 9.3 & 9.3 \end{bmatrix}$$

$W = 5$ and SVD $W = [1][5][1]$ and $U = [1][1/25][1]$

Thus, we have $M = \mathcal{C}U\mathcal{R}$ as

$$M \approx \begin{bmatrix} 1.5 \\ 0 \\ 7.6 \\ 7.6 \\ 7.6 \end{bmatrix} \begin{bmatrix} \frac{1}{25} \end{bmatrix} \begin{bmatrix} 0 & 1.85 & 9.3 & 9.3 \end{bmatrix} = \begin{bmatrix} 0 & .11 & .59 & .59 \\ 0 & 0 & 0 & 0 \\ 0 & .56 & 2.8 & 2.82.8 \\ 0 & .56 & 2.8 & 2.82.8 \\ 0 & .56 & 2.8 & 2.82.8 \end{bmatrix}$$

Try Rank 2 approximation: Possibly select Columns C_3, C_4 and Rows R_3, R_4 and compute scaled columns, rows, matrices W, U and $\mathcal{C}U\mathcal{R}$

Remarks

1. In case a row/column is picked $\beta > 1$ times, we take only one of its copy in $\mathcal{R}/\mathcal{C}$ and scale the corresponding entries by a factor of $\sqrt{\beta}$
2. $\implies W$ may not be square, but we know how to compute SVDs for rectangular matrices.
3. Columns in $\mathcal{C}$ and rows in $\mathcal{R}$ are from M
4. In $\mathcal{C}U\mathcal{R}$ decomposition, U (of dimension at most $\alpha \times \alpha$) may be dense.
5. Total Space $= \alpha(n+m) + \alpha^2$ (likely to be much less due to the sparsity of M)

Performance Guarantee

Quality Estimate

Let M_k be the best rank k-approximation of M.
Choose $\alpha = \frac{k \log k}{\epsilon^2}$.
The resulting $\mathcal{CUR}$ decomposition satisfies the following:
Frobenius Norm of M and $\mathcal{CUR}$ is at most $(2+\epsilon)$ times the Frobenius Norm of M and M_k, i.e. $||M - \mathcal{CUR}||_F \leq (2+\epsilon)||M - M_k||_F$

Remarks:

1. There are recent works that show that $\alpha = k/\epsilon$ suffices
2. Approximation is by a factor of $1+\epsilon$
3. Running time is faster than that of computing SVDs
4. Randomized Linear Algebra - a new field in theoretical computer science.