Fast Computation of Low Rank Matrix Approximations

-Achlioptas , McSherry , journal of ACM 2007

-Nikhil Rao (for CS 838)

Outline

- Computing low rank approximations
- Linear structure and L2 norms
- How this method helps
- Results and proofs

Computing Low Rank Approximations

- Needed to capture only the “meaningful” dimensions
- Traditionally : SVD and Frobenius norm
- SVD also optimal in L2 norm
- Here : L2 norm

Why L2?

- Data in the form of <*instance , attributes*>
- L2 norm measures linear structure : fixing any attribute linearly influences other attributes
- L2 norm = maximum singular value. So the maximum singular vector yields the dimension which captures maximum structure in the data.

Drawback in existing methods

- Lanczos method and Orthogonal iteration.
- Perform matrix vector multiplications for the most part
- At every multiplication, A has to be moved into memory
- Make this more tractable by

1. Sparsifying A , or
2. Quantizing A

To measure error :

$$\psi_A = \min_{Q \in \{-b,+b\}^{mxn}} \|Q\|_2$$

- gives a measure of the amount of linear structure present in an "A – like" matrix
- Will be used to measure error between using the actual and approximated matrix.

The main result

Theorem 2.1 *Let A be an m X n matrix , and let b be as defined earlier. Let:*

$$\hat{A}_{ij} = +-bw.p.1/2 + -A_{ij}/2b$$

then , with probability at least $1 - 1/(m+n)$

$$\|A - \hat{A}_k\| \leq \|A - A_k\| + 7\psi_A$$

Furedi and Komlos , 1980:

Theorem 2.2 *If F is a random symmetric matrix , such that $F_{ij} = r_{ij}$ such that $E[r_{ij}] = 0, Var[r_{ij}] \leq \sigma^2, r_{ij} \in [-K, K]$. If , for any $\alpha > 1/2$,*

$$K < \sigma\sqrt{m+n}(7\alpha log(m+n))^{-3}$$

then

$$\mathcal{P}(\|F\| > 7/3\sigma\sqrt{m+n}) < (m+n)^{1/2-\alpha} \tag{2}$$

Lemma 2.3

$$\psi_A \geq b/\sqrt{2}\sqrt{m+n} \tag{3}$$

Proof *we use the fact that* $\|A\|_F^2 \leq \|A\|_2^2 min(m,n)$

$$\begin{aligned} \|A\|_2 &\geq \frac{\|A\|_f}{\sqrt{min(m,n)}} \\ &= \sqrt{\frac{mn}{min(m,n)}}b \\ &= b\sqrt{max(m,n)} \quad \geq b\sqrt{\frac{m+n}{2}} \end{aligned}$$

Lemma 2.4 $\|\hat{A} - \hat{A}_k\| \leq \|A - A_k\| + \|\hat{A} - A\|$

Proof *From the minimax characterization of the singular values of* $\hat{A}$ *, we get*

$$
\begin{aligned}
\|\hat{A} - \hat{A}_k\| &= \min_{dim Y = k} \max_{x \perp Y} \|\hat{A}x\| \\
&= \min_{dim Y = k} \max_{x \perp Y} \|(A + \hat{A} - A)x\| \\
&\leq \min_{dim Y = k} \max_{x \perp Y} \|Ax\| + \min_{dim Y = k} \max_{x \perp Y} \|(\hat{A} - A)x\| \\
&\leq \min_{dim Y = k} \max_{x \perp Y} \|Ax\| + \max_{x} \|(\hat{A} - A)x\| \\
&= \|A - A_k\| + \|\hat{A} - A\| \quad \blacksquare
\end{aligned}
$$

Theorem 2.5 Let A be an m x n matrix , and let $\hat{A}$ be a random matrix such that $E[\hat{A}_{ij}] = A_{ij}$ and $var(\hat{A}_{ij}) \leq (\sigma b)^2$. Also suppose $|A_{ij} - \hat{A}_{ij}| \leq \sigma b\sqrt{m+n}(7\alpha log(m+n))^{-3}$. Then

$$\|A - \hat{A}_k\| \leq \|A - A_k\| + 7\sigma\psi_A \tag{4}$$

Proof

$$\begin{aligned} \|A - \hat{A}_k\| &\leq \|A - \hat{A}\| + \|\hat{A} - \hat{A}_k\| \\ &\leq \|A - \hat{A}\| + \|A - A_k\| + \|\hat{A} - A\| \{from\ lemma\} \\ &= \|A - A_k\| + 2\|\hat{A} - A\| \end{aligned}$$

Proof contd…

$A - \hat{A}$ *satisfies the conditions for theorem 2.2. Hence , with high probability*

$$\begin{aligned} 2\|A - \hat{A}\| &\leq 2 \times \frac{7}{3}\sigma b\sqrt{m+n} \\ &= 7\sigma\frac{2}{3}b\sqrt{m+n} \\ &\leq 7\sigma b\sqrt{\frac{m+n}{2}} \\ &\leq 7\sigma\psi_A \end{aligned}$$

And finally…

- By making the variance of the random entries in the matrix to be unity , we get the result in theorem 2.1
- The result of the sparse approximation of A can be obtained by letting the variance to be a a function of `s' , a parameter in the sparsification result.

Thank You