Incremental Singular Value Decomposition Algorithms for Highly Scalable Recommender Systems (Sarwar et al)

Presented by Sameer Saproo

Recommender Systems

- Apply Knowledge Discovery in Databases (KDD) to make personalized product recommendations during live customer interaction
- Offline Vs Online
- Not Google!

CF-based Recommender systems

- Suggest new products or suggest utility of a certain product for a particular customer, based on customer's previous liking and the opinions of other like-minded customers

	Matrix	Pi	AI
Alice	5	3	x
Bob	x	3	5
Carol	5	x	x

CHALLENGES

- Quality of Recommendation (Q)
- Scalability of CF Algorithms (S)

$$Q \propto \frac{1}{S}$$

- SVD based Latent Semantic Indexing presents an approach to CF based recommendations, but stumbles in Scalability
- The paper produces an algorithm for improving scalability for SVD based CF by sacrificing accuracy a little.

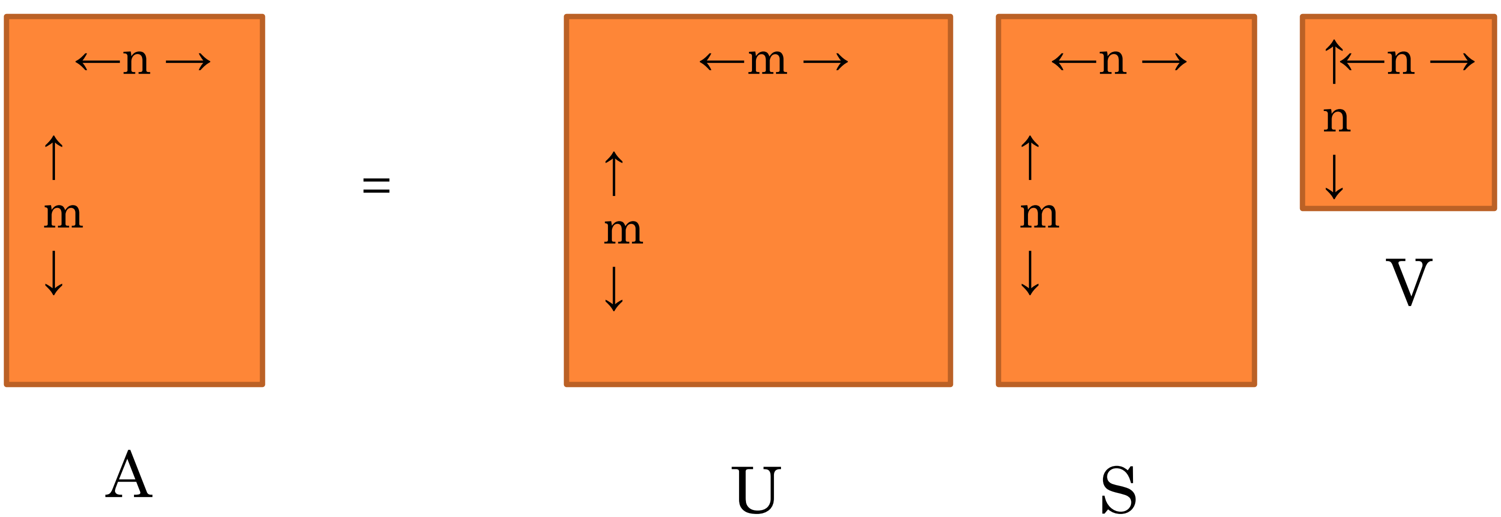
In Nutshell

- Problem
  - The matrix factorization step in SVD is computationally very expensive
- Solution
  - Have a small pre-computed SVD model, and build upon this model incrementally using inexpensive techniques

SINGULAR VALUE DECOMPOSITION

- Matrix factorization technique for producing low-rank approximations

$$SVD(A) = U \times S \times V^T$$

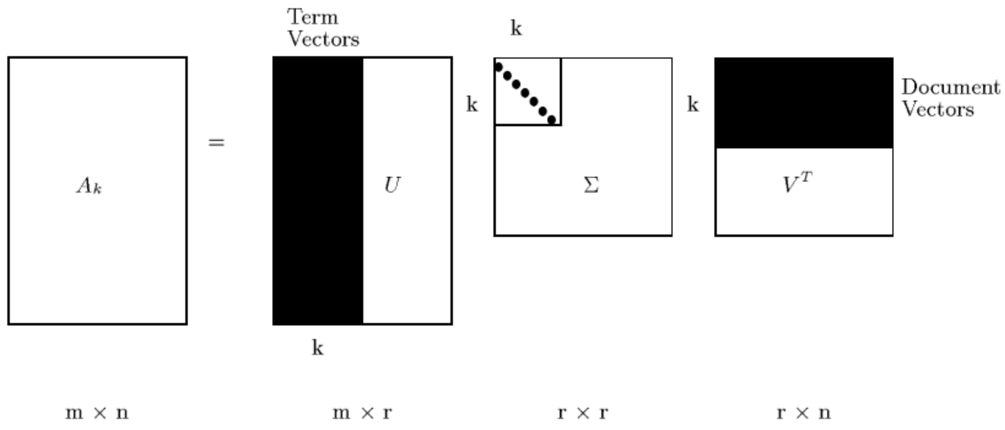


Low Rank Approximation (USV^T)

- U and V are orthogonal matrices and S is a diagonal matrix
- S has r non-zero entries for a rank r matrix A.
- Diagonal Entries (S_1, S_2, S_3, S_4,…, S_r) have the property that $S_1 \geq S_2 \geq S_3 \geq S \geq \ldots \geq S_r$
- SVD provides best *low-rank* linear approximation of the original matrix A i.e. if

$A_k = U_k . S_k . V_k^T$ is a rank - k matrix whi ch is the closest approximat ion of A. More Specifical ly, A_k minimizes the Frobenius Norm $\|A - A_k\|_F$, where a Frobenius Norm $\|A\|_F$ is defined as $\sqrt[2]{\sum_{ij} |a_{ij}|^2}$

CONTD.



- A low-rank approximation of the original space is better than the original space as small singular values which introduce noise in customer-product matrix are filtered out.
- SVD produces uncorrelated eigenvectors, and each customer/product is represented by its own eigenvector.
- This dimensionality reduction helps customers with similar taste to be mapped into space represented by same eigenvectors.

Prediction Generation

- Formally,

$$P_{i,j} = \bar{r}_i + \left(U_k . \sqrt{S_k}^T (i)\right) . \left(\sqrt{S_k}^T . V_k (j)\right)$$

where,

$P_{i,j}$ is the prediction for ith customer and jth product .

$\bar{r}_i$ is the row average.

We calculate the cosine similariti es between between m pseudo -

customers $U_k . \sqrt{S_k}^T$ and n pseudo - products $\sqrt{S_k}^T . V_k$

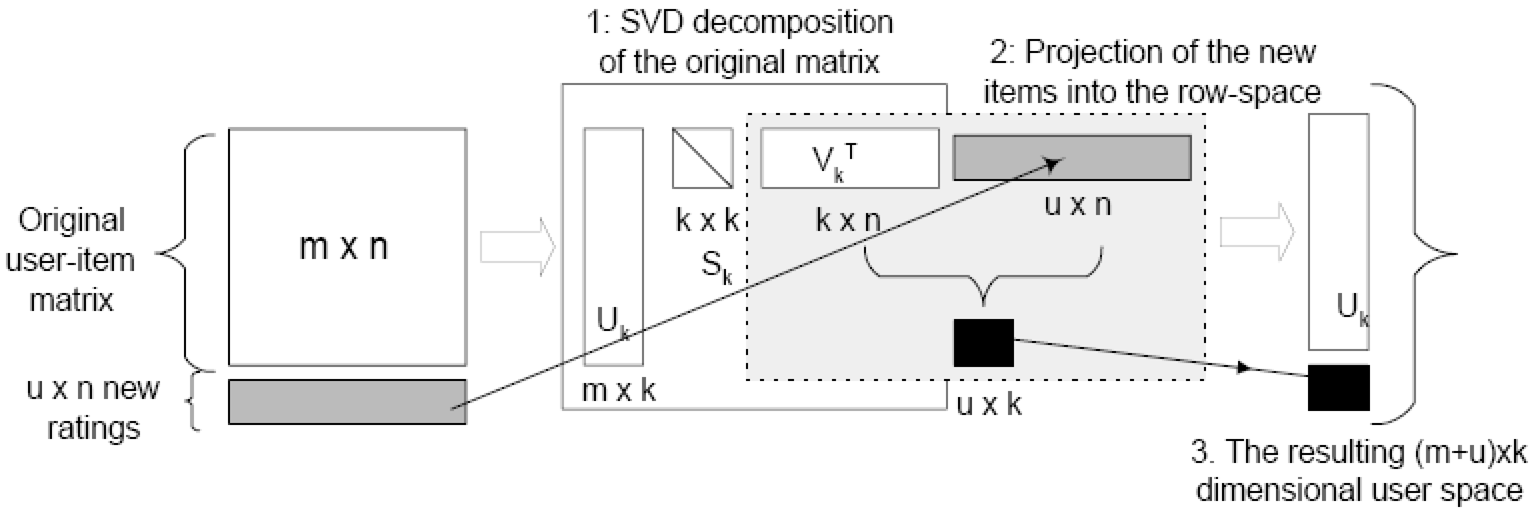
Challenges of Dimentionality reduction

- Offline Step
  - Also known as Model Building
  - User-user similarity computation and neighborhood formation i.e. SVD decomposition
  - Time consuming and infrequent
  - $O(m^3)$ for m x n matrix A
- Online Step
  - Also known as Execution step
  - Actual prediction generation
  - O(1)

Incremental SVD Algorithms

- Borrowed from the LSI world to handle dynamic databases
- Projection of additional users provides good approximation to the complete model
- Authors build a suitably sized model first and then use projections to incrementally build on that
- Errors induced as the space is not orthogonal

Folding-In



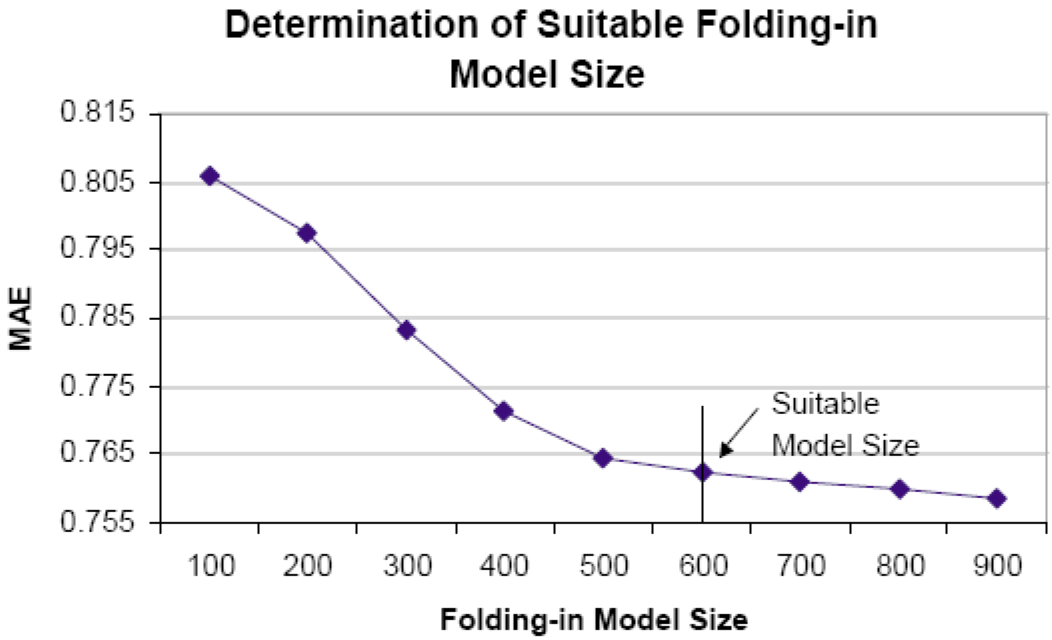
As depicted in the paper	*Found in Reference [1]*
New user vector N_u be $t \times 1$ $P = U_k \times U_k \times N_u$ Append k-dimensional vector $U_k^T . N_u$ as a new column of the $k \times d$ matrix $S_k . V_k^T$	t is $1 \times n$ user vector its projection on the span of current product vectors (columns of V_k) $\hat{t} = t V_k \Sigma_k^{-1}$ Appended to columns of U_k

Experimental Evaluation

- Dataset : www.movielens.umn.edu
- About 100,000 ratings
- User – Movie matrix : 943 users and 1682 movies
- Training – Test ratio : 80%
- Evaluation Metric
  - Mean Absolute Error (MAE) = $\frac{\sum_{i=1}^{N} |p_i - q_i|}{N}$
  - $< p_i - q_i >$ is a ratings – prediction pair

Model Size

Optimal reduced Rank k=14 was found empirically

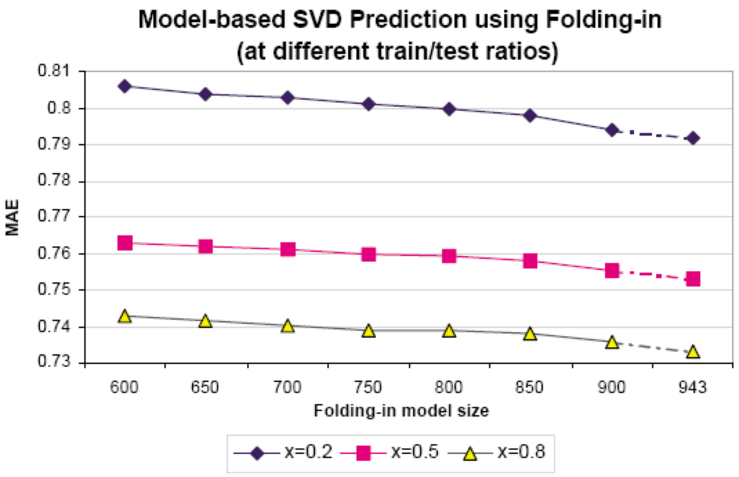


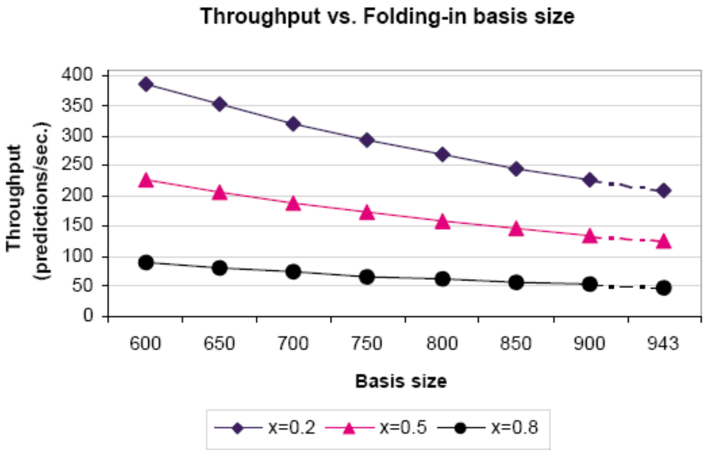
(943 – Model size) is projected using folding-in

Results

Quality

Performance





For Model size of 600, quality loss was 1.22% whereas performance increase was 81.63%


MERRIE MELODIES
REG. US PAT. OFF.
"That's all Folks!"
A WARNER BROS. CARTOON
DUBBED VERSION ©1995 TURNER ENTERTAINMENT CO.
MUSIC ©1995 WARNER BROS. ℗1995 WARNER BROS.
ALL LOGOS AND CHARACTERS ARE TRADEMARKS OF WARNER BROS.