

Goal

Given 100,000,000 *known* ratings of movies by users, and 3,000,000 *unknown* ratings of movies by users, predict what these unknown ratings are *as accurately as possible*.



Ryan Batterman
RBM

Joseph Choi
Enhanced timeSVD++
Temporal Conditional Factored RBM
Poster

Joshua Dale
k-Nearest Neighbors

Fundamental Idea of Singular Value Decomposition

Each user can be represented as a set of preferences for different features of movies, and each movie can be represented as a set of how much of each feature that movie has.

By taking the dot product of these two vectors, one can predict what a user will rate a movie.

Enhanced timeSVD++

$$\hat{r}_{uit} = b_{uit} + \sum_k \left[(q_{ik} + q_{iBin(t)k} + q_{iftk})^T \left(p_{uk} + \alpha_{uk} dev_u(t) + p_{utk} + \frac{1}{\sqrt{N(u)}} \sum_{j \in N(u)} (y_{jk} + y_{jBin(t)k}) \right) \right]$$

$$b_{uit} = \mu + b_u + \alpha_u dev_u(t) + b_{ut} + (b_i + b_{iBin(t)}) (c_u + c_{ut}) + b_{ift}$$

Temporal Conditional Factored RBM

$$\hat{r}_{uit} = \frac{\sum_k k f(k)}{\sum_k f(k)}$$

$$f(k) = \exp \left\{ b_u^k + b_{ut}^k + b_i^k + b_{ift}^k + \sum_{j=1}^F \sum_{c=1}^C \left[A_{ic}^k B_{cj} \sigma \left(b_j + \sum_{i=1}^m \sum_{k=1}^K \sum_{c=1}^C \delta_{r_ik} A_{ic}^k B_{cj} + \sum_{i=1}^M r_i D_{ij} \right) \right] \right\}$$

The Power of Blending

$$\hat{Q} = R(R^T R)^{-1} R^T Q$$

$$(R^T Q)_i = \sum_j R_{ij} Q_j$$

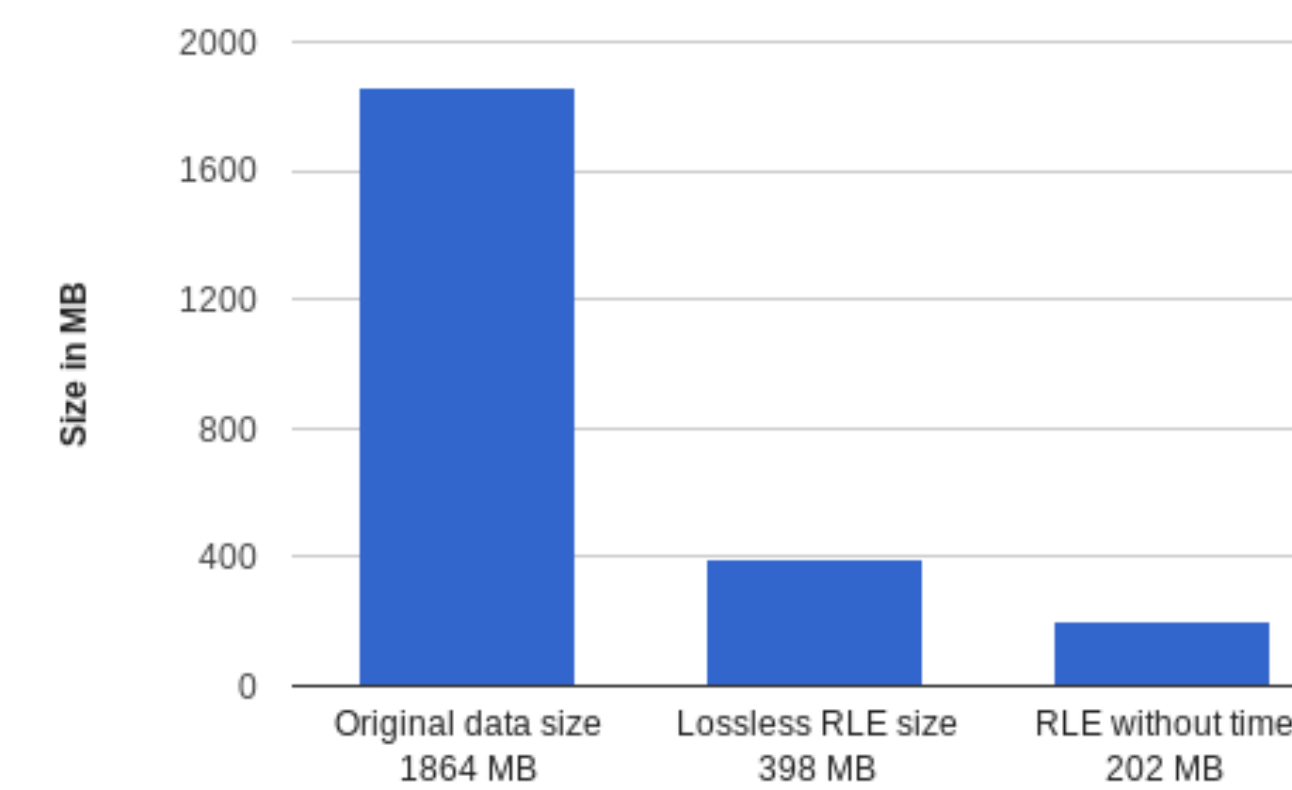
$$RMSE_{quiz} \Rightarrow \sum_j (R_{ij} - Q_j)^2 = \sum_j R_{ij}^2 - 2 \sum_j R_{ij} Q_j + \sum_j Q_j^2$$

Quiz RMSE **9.1234%**

Ingredients:
10 Joseph's **enhanced timeSVD++**
50 Joseph's **enhanced timeSVD++**, **overfit**
5 Joseph's **temporal conditional RBM**, **overfit**
5 Joseph's **temporal conditional factored RBM**, **overfit**

Fast Loading

0.04 seconds to load entire dataset
0.08 seconds to load entire dataset + time



Fast Converging

	SVD	SVD++	timeSVD++
Time per iteration	1.4 seconds	2.4 seconds	3.4 seconds

All SVD variants benchmarked at 40 features.

	RBM	CRBM	TCRBM	TCFRBM
Time per iter.	3.8 seconds	4.4 seconds	4.8 seconds	6.9 seconds

All RBM variants benchmarked at 10 features.

