Exact and Stable Recovery of Pairwise Interaction Tensor

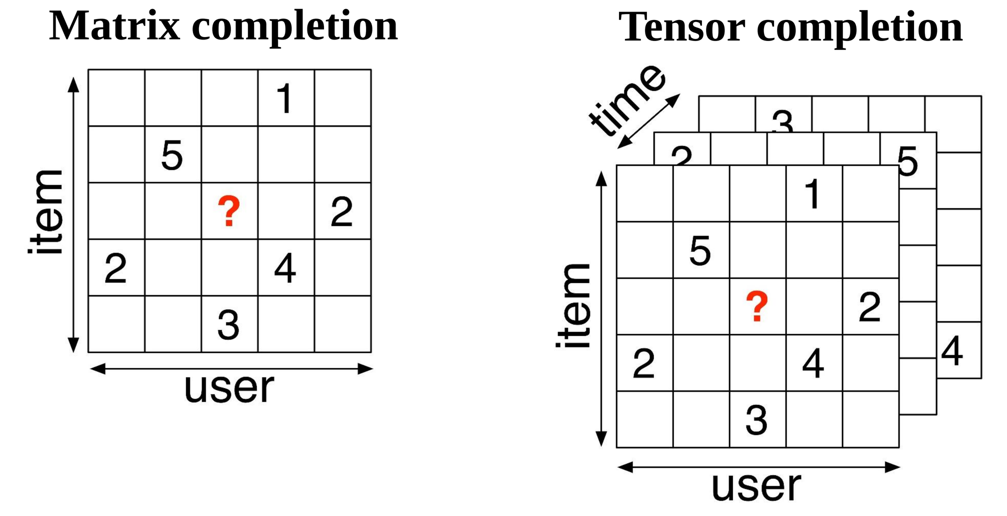
Neural Information Processing Systems Foundation

[1]Shouyuan Chen, [1]Michael R. Lyu, [1]Irwin King, [2]Zenglin Xu

[1]The Chinese University of Hong Kong [2]Purdue University

香港中文大學 The Chinese University of Hong Kong

Tensor Completion



Recover Pairwise Interaction Tensor

Object	Decomposition	Recovery
rank-k matrix $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$	$M_{ij} = \langle u_i, v_j \rangle$	matrix completion guaranteed recovery of **M** from $O(nk \log^2(n))$ observations
rank-k tensor $\mathbf{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$	$T_{ijk} = \langle u_i, v_j, w_k \rangle$	???
pairwise interaction tensor $\mathbf{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$	$T_{ijk} = \langle u_i^{(a)}, v_j^{(a)} \rangle + \langle u_j^{(b)}, v_k^{(b)} \rangle + \langle u_k^{(c)}, v_i^{(c)} \rangle$	this paper: guaranteed recovery of **T** from $O(nk \log^2(n))$ observations.

Previous Pairwise Interaction Tensor

Tag recommendation [1] Sequential analysis of purchase data [2]

- model data using pairwise interaction tensor instead of general low rank tensors.
- faster/more accurate/achieves state of the art performance.

Factorization Machine [3]

- extend to higher order tensors.

Existing learning algorithms are prone to local optimal issues

- recovered tensor can be very different from its true value!

References

[1] Rendle, Steffen, and Lars Schmidt-Thieme. "Pairwise interaction tensor factorization for personalized tag recommendation." WSDM 2010.

[2] Rendle, Steffen, Christoph Freudenthaler, and Lars Schmidt-Thieme. "Factorizing personalized Markov chains for next-basket recommendation." WWW, 2010.

[3] Rendle, Steffen. "Factorization machines with libFM." TIST 2012.

Matrix Formulation

Original formulation: $T_{ijk} = \langle u_i^{(a)}, v_j^{(a)} \rangle + \langle u_j^{(b)}, v_k^{(b)} \rangle + \langle u_k^{(c)}, v_i^{(c)} \rangle$

Equivalent formulation: $T_{ijk} = A_{ij} + B_{jk} + C_{ki}$

for all $i, j, k \in [n_1] \times [n_2] \times [n_3]$

Denote $\mathbf{T} = \mathrm{Pair}(\mathbf{A}, \mathbf{B}, \mathbf{C})$

Recovery Problem

Given: partial observations Ω of a pairwise interaction tensor T

Goal: recover matrices A, B, C and therefore T.

Result: Exact Recovery

When all observations of **T** are **exact** and **noiseless**, we can **exactly recover** the pairwise interaction tensor from a subset of observations.

Solve a weighted trace norm minimization problem:

$$\min_{\mathbf{X} \in S_A, \mathbf{Y} \in S_B, \mathbf{Z} \in S_C} \sqrt{n_3}\|\mathbf{X}\|_* + \sqrt{n_1}\|\mathbf{Y}\|_* + \sqrt{n_2}\|\mathbf{Z}\|_*$$
$$s.t.\ X_{ij} + Y_{jk} + Z_{ki} = T_{ijk}, \qquad \forall (i,j,k) \in \Omega.$$

Theorem: Under mild assumptions (see below), if the number of observations is larger than $\boldsymbol{O}\left(\boldsymbol{n_3 r \log^2(n_3)}\right)$, then, with high probability, the minimizing solution of the above objective satisfies $\mathbf{A} = \mathbf{X}, \mathbf{B} = \mathbf{Y}$ and $\mathbf{C} = \mathbf{Z}$ and therefore exactly recovers pairwise interaction tensor **T**.

Result: Stable Recovery

When the observations are **noisy**, the performance of recovery is **accurate**.

Let $\hat{T}$ be the tensor perturbed by noise. Assume $\left\|P_\Omega(\hat{T} - T)\right\|_F \le \epsilon_1$.

Solve a weighted trace norm minimization problem:

$$\min_{\mathbf{X} \in S_A, \mathbf{Y} \in S_B, \mathbf{Z} \in S_C} \sqrt{n_3}\|\mathbf{X}\|_* + \sqrt{n_1}\|\mathbf{Y}\|_* + \sqrt{n_2}\|\mathbf{Z}\|_*$$
$$s.t. \left\|P_\Omega\big(\mathrm{Pair}(\mathbf{X}, \mathbf{Y}, \mathbf{Z})\big) - P_\Omega(\hat{\mathbf{T}})\right\|_F \le \epsilon_2.$$

Theorem: Under same assumptions, if the number of observations is larger than $\boldsymbol{O}\left(\boldsymbol{n_3 r \log^2(n_3)}\right)$, then the minimizing solution of the above objective satisfies

$$\|\mathrm{Pair}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) - \mathbf{T}\|_* \le \tilde{O}\left(r^{0.5} n^{1.5}(\epsilon_1 + \epsilon_2)\right).$$

Conditions of Recovery

Incoherence.

- Matrix completion is a special case of our problem (e.g. recover Pair(**A**,0,0)).
- Incoherence is an essential requirement of matrix completion.
- Our results inherit the incoherence conditions, i.e. **both theorem require that A,B,C are incoherent**.

Uniqueness.

- A pairwise interaction tensor **T** has infinite many equivalent matrix representations.
- **Unique representation**: for any pairwise interaction tensor $\mathbf{T} = \mathrm{Pair}(\mathbf{A}', \mathbf{B}', \mathbf{C}')$, there exists unique $\mathbf{A} \in S_A, \mathbf{B} \in S_B, \mathbf{C} \in S_C$ such that $\mathrm{Pair}(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \mathrm{Pair}(\mathbf{A}', \mathbf{B}', \mathbf{C}')$
- Our results assume that $\mathbf{A} \in S_A, \mathbf{B} \in S_B, \mathbf{C} \in S_C$.
- Construction of S_A, S_B, S_C is related to the "bias" component.

Optimization Algorithm

We use SVT to solve the trace norm minimization problem.

Iterate between Step (1) and Step (2) ..

Step (1)

$$\mathbf{X}^k = \mathrm{shrink}_A\left(P_{\Omega_A}^*(\mathbf{y}^{k-1}), \tau\right)$$
$$\mathbf{Y}^k = \mathrm{shrink}_B\left(P_{\Omega_B}^*(\mathbf{y}^{k-1}), \tau\right)$$
$$\mathbf{Z}^k = \mathrm{shrink}_C\left(P_{\Omega_C}^*(\mathbf{y}^{k-1}), \tau\right)$$

Step (2) (for exact recovery)

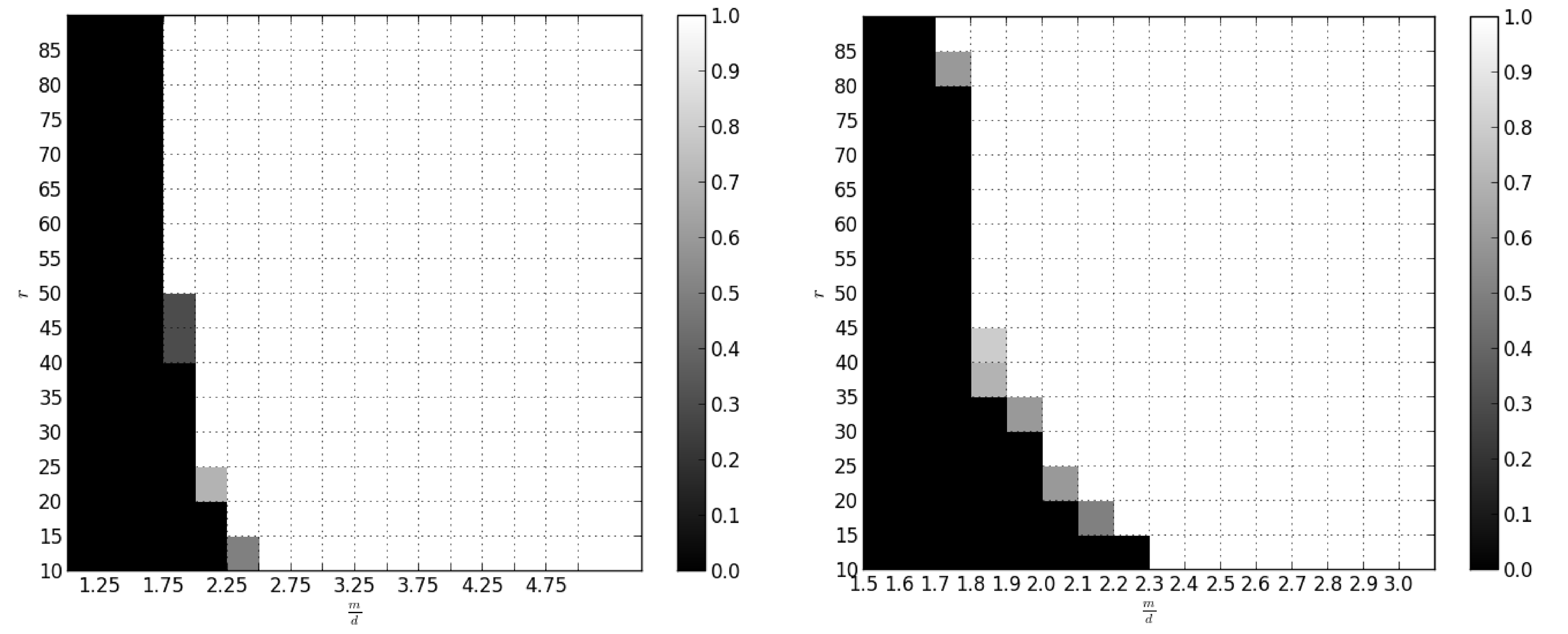
$$\boldsymbol{e}^k = P_\Omega(\mathbf{T}) - P_\Omega\left(\mathrm{Pair}\left(\frac{\mathbf{X}^k}{\sqrt{n_3}}, \frac{\mathbf{Y}^k}{\sqrt{n_1}}, \frac{\mathbf{Z}^k}{\sqrt{n_2}}\right)\right)$$
$$\boldsymbol{y}^k = \boldsymbol{y}^{k-1} + \delta \boldsymbol{e}^k.$$

The **shrinkage operator** shrink_ω is defined as

$$\mathrm{shrink}_\omega(\mathbf{M}, \tau) \stackrel{\text{def}}{=} \arg\min_{\widetilde{\mathbf{M}} \in S_\omega} \frac{1}{2}\left\|\mathbf{M} - \widetilde{\mathbf{M}}\right\|_F + \tau\left\|\widetilde{\mathbf{M}}\right\|_*$$

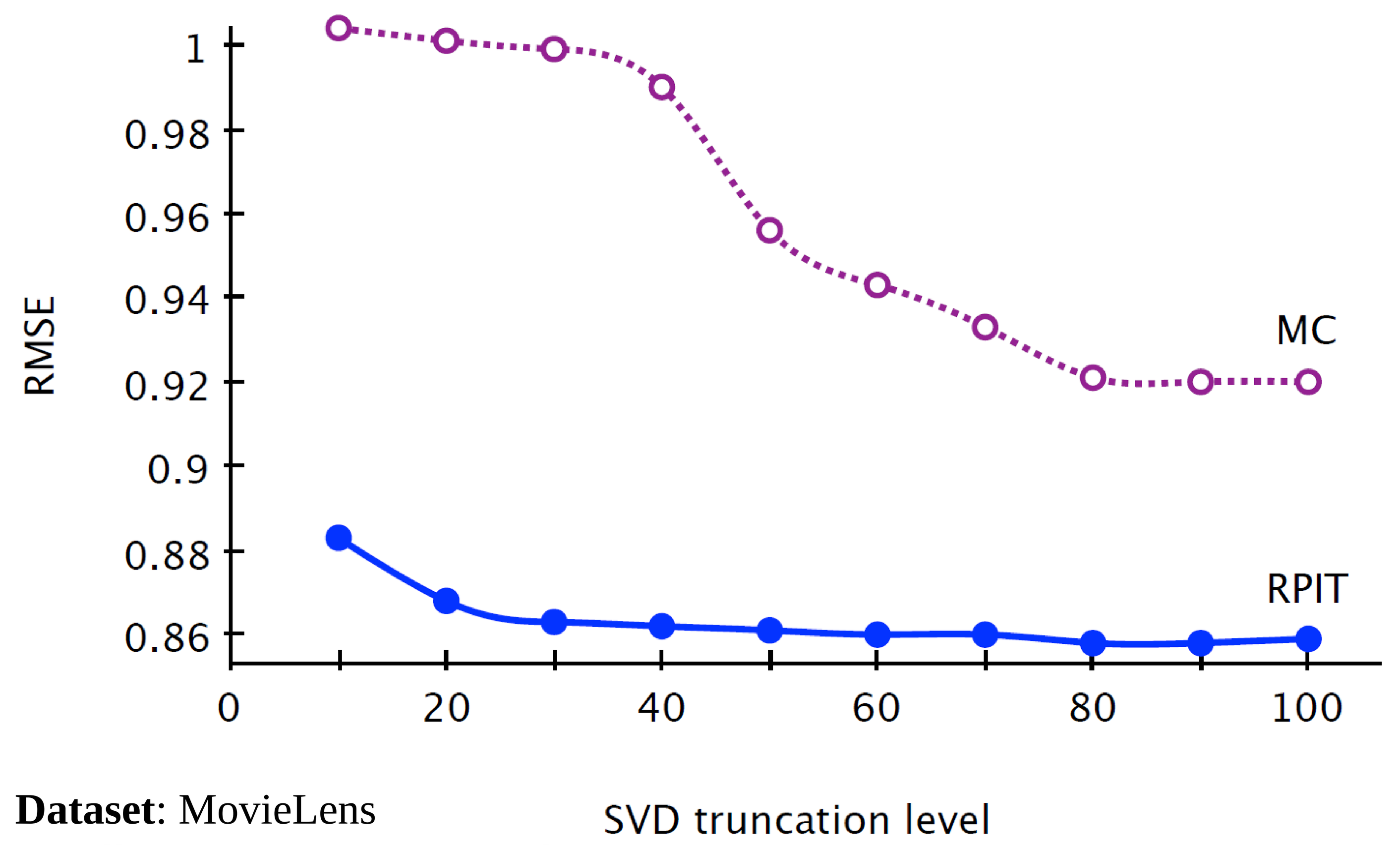
The shrinkage operators can be computed efficiently using **SVD**.

Phase Transition of Exact Recovery



- The x-axis is the ratio between the number of observations m and the degree of freedom.
- The y-axis is the rank r of the **synthetic** matrices **A,B,C**.
- The color of each grid indicates the empirical success rate.

Temporal Collaborative Filtering



- **Dataset**: MovieLens
  - All ratings are timestamped.
- **Model**: Tensor N*M*T, N: number of users, M: number of movies, T: number of different months.
  - size: 6040*3706*36, observations: 1M.
- **Algorithms**:
  - MC: Matrix completion, which does not use timestamp information.
  - RIPT: Our algorithm, which uses timestamp information, achieves RMSE of 0.861