

Part II: Rank Aggregation via Nuclear Norm Minimization

David F. Gleich, Lek-Heng Lim

NIPS Workshop on Advances in Ranking

December 11, 2009

Noiseless Recovery Problem

Problem (Lowest rank skew-symmetric matrix recovery)

Given $\mathcal{A} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^m$ linear operator and $b \in \mathbb{R}^m$.

$$\min\{\text{rank}(X) \mid X^T = -X, \mathcal{A}(X) = b\}.$$

- Convex relaxation along the lines of compressive sensing?

$$\|\cdot\|_1 \approx \|\cdot\|_0, \quad \|\cdot\|_* \approx \text{rank}.$$

- Ky Fan/nuclear/Schatten/trace norm,

$$\|A\|_* = \sum_{i=1}^{\text{rank}(A)} \sigma_i(A).$$

- Nuclear norm relaxation

$$\min\{\|X\|_* \mid X^T = -X, \mathcal{A}(X) = b\}$$

[Recht-Fazel-Parrilo; 09], [Candès-Tao; 09], [Meka-Jain-Dhillon; 09].

- Problem: $\mathcal{A}(X) = b$ may not have a rank-2 solution.

Noisy Recovery Problem

- Replace $\mathcal{A}(X) = b$ by $\mathcal{A}(X) \approx b$.
- Various formulations

BPDN:

$$\min\{\|X\|_* \mid X^T = -X, \|\mathcal{A}(X) - b\|_2 \leq \sigma\}$$

LASSO:

$$\min\{\|\mathcal{A}(X) - b\|_2 \mid X^T = -X, \|X\|_* \leq \tau\}$$

QP:

$$\min\{\|\mathcal{A}(X) - b\|_2^2 + \lambda\|X\|_* \mid X^T = -X\}$$

DS:

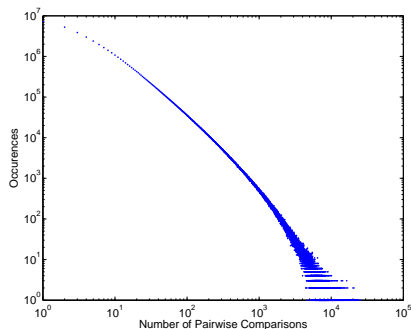
$$\min\{\|X\|_* \mid X^T = -X, \|\mathcal{A}^*(\mathcal{A}(X) - b)\|_{2,2} \leq \mu\}$$

- Want: simplest non-trivial skew-symmetric X , i.e. $\text{rank}(X) = 2$.
- So $X_{ij} = (s_i - s_j)$ for $i, j = 1, \dots, n$ and $s = [s_1, \dots, s_n]$ consistent aggregate score.

Netflix and MovieLens Pairwise

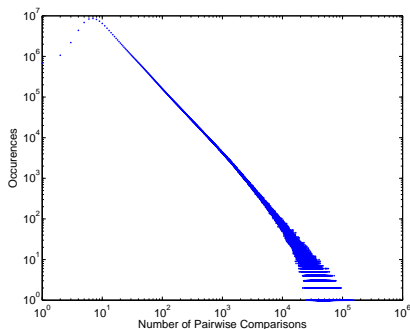
Usually the pairwise ranking matrices Y are almost dense. However, many of the entries only have a few comparisons. We remove entries with less than 30 comparisons.

MovieLens 10M



85.49% to 18.50%

Netflix



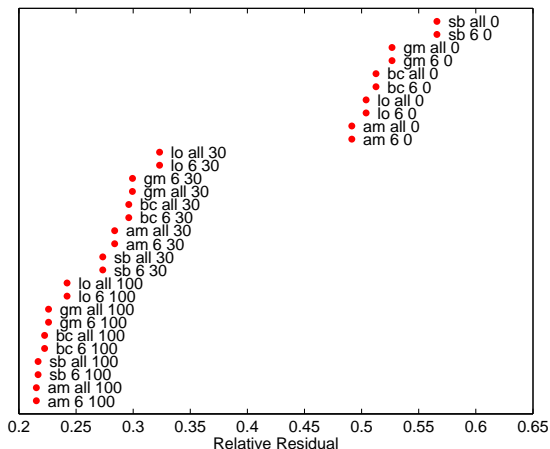
99.77% to 34.67%

This means that Y with infrequent elements removed is a nice candidate for matrix completion. (Also, note the hump at 10 ratings for Netflix!)

Residual Results from Netflix Rankings

Experiment

Solve LASSO problem with the Meka, Jain, Dhillon SVP solver for rank-2 matrices with Y computed from all models.



Linear	am
Log-linear	gm
Linear Prob.	bc
Bradley-Terry	lo

RMSE values are smaller 0.05 to 0.15, $\sim 100M$ terms

bc 6 100 \rightarrow Linear Prob. model, all users with at least 6 ratings, at least 100 pairwise comparisons.

Top Movies from Netflix

Linear Full	Linear 30 Pairwise	Bradley-Terry Full
Greatest Story Ever ...	LOTR III: Return ...	LOTR III: Return ...
Terminator 3	LOTR I: The Fellowship ...	LOTR II: The Two ...
Michael Flatley	LOTR II: The Two ...	LOTR I: The Fellowship ...
Hannibal [Bonus]	Star Wars VI: Return ...	Star Wars V: Empire ...
Donnie Darko [Bonus]	Star Wars V: Empire ...	Raiders of the Lost Arc
Timothy Leary's ...	Star Wars IV: A New Hope	Star Wars IV: A New Hope
In Country	LOTR III: Return ...	Shawshank Redemption
Bad Boys II [Bonus]	Raiders of the Lost Arc	Star Wars VI: Return ...
Cast Away [Bonus]	The Godfather	LOTR III: Return ...
Star Wars: Ewok ...	Saving Private Ryan	The Godfather

LOTR III shows up twice because of the two DVD editions.

Full model has many “Bonus” discs that Netflix rents. These are items enjoyed by only a few people.

Rank 2 vs. Rank 4

The residual does not greatly change from rank 2 approximations to rank 4 approximations.

Model	Rank	RMSE
Linear (6,100)	2	0.174
Linear (6,100)	4	0.154

Preliminary Conclusions

Summary

- The initial results are promising ...
- It is sensitive to the choice of model for Y .

Future work

- Comparison against Hodge theory approach.
- Validating rank-2 assumption (Owen and Perry's bi-cross validation approach)
- Sensitivity to removing more entries.