Collaborative Filtering and Optimal Transport for Recommender Systems

Ellington Kirby, Mehdi Inane, and Jules Merigot

Université Paris Dauphine

October 16, 2023

Introduction

- Predict user ratings using sparse data
- Implemented Matrix Factorization with Gradient Descent and Alternating Least Squares
- Implemented Inverse Optimal Transport with Two Algorithms: Inverse Optimal Transport (IOT) and Regularized Inverse Optimal Transport (RIOT)

Results

Method	RMSE Train	RMSE Test	Time
Gradient Descent	0.89	0.98	16.46 s
ALS	0.77	0.88	25.62 s
IOT	1.16	1.15	2.1 s
RIOT	2.45	2.60	534 s

Table 1: RMSE and Time Results of our Four Methods

Alternating Least Squares

- Random initialization of user matrix X and item matrix Y [1]
- Compute:

$$x_u = \left(\sum_{r_{ui} \in r_{u*}} y_i y_i^T + \lambda I_k \right)^{-1} \sum_{r_{ui} \in r_{u*}} r_{ui} y_i$$

$$y_i = \left(\sum_{r_{ui} \in r_{*i}} x_u x_u^T + \lambda I_k \right)^{-1} \sum_{r_{ui} \in r_{*i}} r_{ui} x_u$$

- Compute error difference after each iteration using binary matrix of observed values
- Rounding up or down VS np.clip()

Alternating Least Squares

- Implemented grid search to tune parameters for better performance
- Found best to be:
  - rank: $k = 1$
  - iterations: 125
  - regularization: $\lambda = 0.34$
  - error tolerance: 1×10^{-6}

ALS Optimization	RMSE Train	RMSE Test	Time
Before Grid Search	0.786	0.886	21.63 s
After Grid Search	0.766	0.884	27.9 s

Table 2: RMSE and Time Results for ALS before and after grid search implementation

Optimal Transport

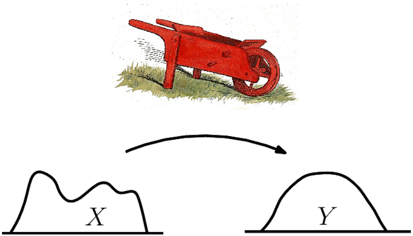


Figure 1: Transporting from one space to another [1]

[1] https://sbl.inria.fr/doc/Earth_mover_distance-user-manual.html

Inverse Optimal Transport

- Set $C = f(U^T AV)$
- The inverse optimization problem :

$$\min_C -\sum_{i=1}^{m}\sum_{j=1}^{n} \hat{\pi_{ij}} \log \pi_{ij} = \min_C KL(\hat{\pi}, \pi) \tag{1}$$

- The alternating min-max problem

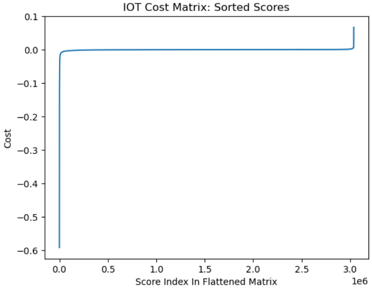
$$\min_{A,\mu,\nu} \max_{z,w} -\sum_{i=1}^{m}\sum_{j=1}^{n} \hat{\pi_{ij}} \log \pi_{ij} + \lambda(\hat{\mu}, \hat{\nu}, z, w, \mu, \nu, U, V) \tag{2}$$

Where $\hat{\pi}$ is the normalized ratings matrix, and $\hat{\mu}, \hat{\nu}$ are the empirical marginal distributions obtained from the empirical matrix.

Methodology

- Interpret Cost matrix or Compute Optimal Transport Plan
  Cost Matrix Interpretation: Lower RMSE. Could be because of supply limitations.
- Non-Negative Matrix Factorization or User-User/Item-Item Dissimilarity Matrices
  Found NMF to have better results and smaller matrices, but less interpretable.
- Type of Regularization used in IOT L1 vs L2 vs Strictly Positive Constraints.

Impact of Methodological Choices



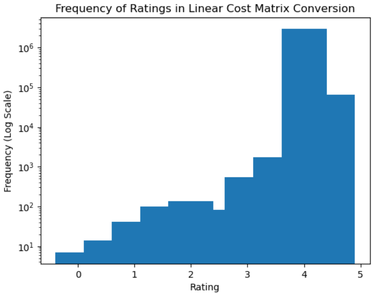
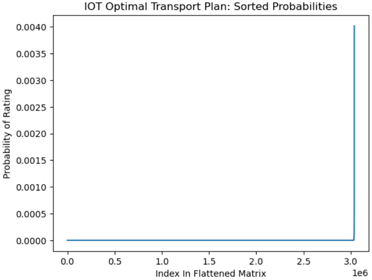


Figure 2: Distribution of Scores for IOT Cost Matrix Interpretation. RMSE of 1.15. Hypothesis: the density of 4 star ratings might be the culprit. Note that a similar effect was observed with RIOT, but not centered around a rating of 4.

Impact of Methodological Choices



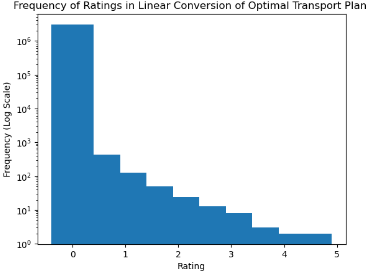


Figure 3: Distribution of Scores after running OT using Earth Movers Distance using Cost Matrix learned via IOT, RMSE of 3.66. Note the high distribution of zeros.

Conclusion: Sparsity causes many issues with OT as a method. Further, we were unable to come up with a valid interpretation of element probabilities or costs as scores; our linear estimation is not sufficient. This is the an avenue for future work.

Publication and References

1. Reza, B. (2015). Notes on Alternating Least Squares. CME 323: Distributed Algorithms and Optimization, Stanford University. Retrieved from https://stanford.edu/ rezab/classes/cme323/S15/notes/lec14.pdf
2. Li, R., Ye, X., Zhou, H., Zha, H. (2019). Learning to match via inverse optimal transport. Journal of machine learning research, 20.
3. Ma, S., Sun, H., Ye, X., Zha, H., Zhou, H. (2020). Learning cost functions for optimal transport. arXiv preprint arXiv:2002.09650.