Recommender System using Matrix Factorization

Jia-Wei Liao, Kuok-Tong Ng, Yi-Cheng Hung
Advisor: Wen-Wei Lin

Department of Applied Mathematics
National Yang Ming Chiao Tung University

December 29, 2021

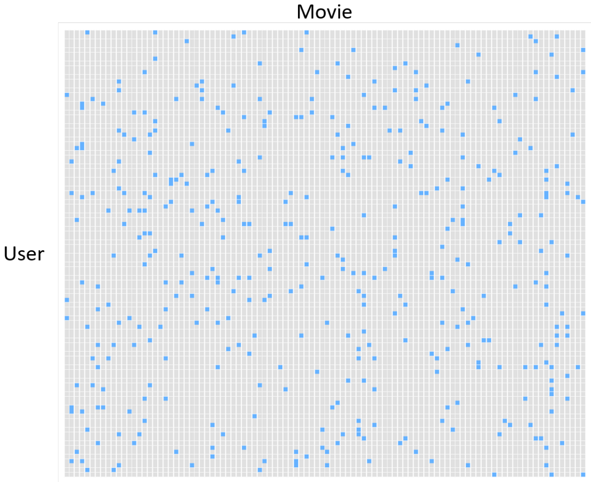
Outline

Introduction

(User, Movie)	Rating
(1, 5)	4
(1, 18)	1
(1, 32)	3
(1, 44)	2
(2, 22)	5
(2, 90)	2
(3, 49)	3
(3, 56)	4
(3, 70)	5
(3, 94)	1
$\vdots$	$\vdots$

Introduction

(User, Movie)	Rating
(1, 5)	4
(1, 18)	1
(1, 32)	3
(1, 44)	2
(2, 22)	5
(2, 90)	2
(3, 49)	3
(3, 56)	4
(3, 70)	5
(3, 94)	1
⋮	⋮



Idea

Action	0.2	1.2	1.8	1.8
Romance	1.68	0.5	0.1	0.3
Science Fiction	0.2	1.8	0.6	1.6
Fantasy	0.48	0.8	0.4	0.4

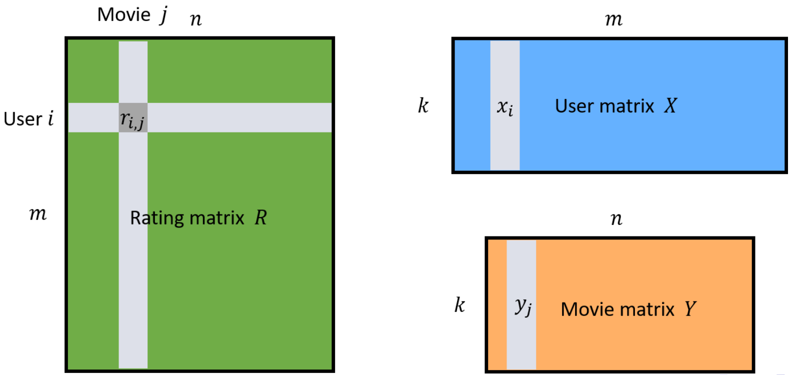
	A.	R.	S.	F.
Sam	0.8	1.6	1.8	1.2
Jack	1.1	0.3	1.2	1.6
Andy	1.3	0.5	1.6	0.8

5.07

Matrix Factorization

Given $R \in \mathbb{R}^{m \times n}$ is rating matrix, where m is number of user, and n is number of movie. Our goal is going to find the feature vecor of user x_i and feature vector of movie y_j such that

$$r_{i,j} \approx x_i^\top y_j, \text{ for all } i, j$$



Matrix Factorization

Problem

$$\min_{X,Y} \|R - X^\top Y\|_F^2 + \lambda \left(\|X\|_F^2 + \|Y\|_F^2\right)$$

where

- $R \in \mathbb{R}^{m\times n}$, $X \in \mathbb{R}^{k\times m}$ and $Y \in \mathbb{R}^{k\times n}$
- $\lambda > 0$ is a parameter

Matrix Factorization

Problem

$$\min_{X,Y} \|R - X^\top Y\|_F^2 + \lambda \left(\|X\|_F^2 + \|Y\|_F^2\right)$$

where

- $R \in \mathbb{R}^{m\times n}$, $X \in \mathbb{R}^{k\times m}$ and $Y \in \mathbb{R}^{k\times n}$
- $\lambda > 0$ is a parameter

Remark

$$\begin{aligned} &\min_{X\in\mathbb{R}^{k\times m}, Y\in\mathbb{R}^{k\times n}} \|R - X^\top Y\|_F^2 + \lambda \left(\|X\|_F^2 + \|Y\|_F^2\right) \\ &= \min_{Z:\operatorname{rank} Z\le k} \|R - Z\|_F^2 + 2\lambda\|Z\|_* \end{aligned}$$

Alternative Least Square (ALS)

ALS method

Given initials $X^{(0)}, Y^{(0)}$, for $t = 0, 1, 2, ...$, we solve the following two sub-problem alternatingly:

$$X^{(t+1)} = \arg\min_{X} \|R - X^\top Y^{(t)}\|_F^2 + \lambda \|X\|_F^2$$

$$Y^{(t+1)} = \arg\min_{Y} \|R - {X^{(t+1)}}^\top Y\|_F^2 + \lambda \|Y\|_F^2$$

We iterate until convergence.

Solving Alternative Least Square

Define

$$F(X) = \|R^\top - {Y^{(t)}}^\top X\|_F^2 + \lambda \|X\|_F^2$$

Then

$$\nabla F(X) = -2Y^{(t)}(R^\top - {Y^{(t)}}^\top X) + 2\lambda X$$

Let $\nabla F(X) = 0$, we have

$$(Y^{(t)}{Y^{(t)}}^\top + \lambda I)X = Y^{(t)} R^\top$$

Therefore, we obtain the solution

$$X^{(t+1)} = (Y^{(t)}{Y^{(t)}}^\top + \lambda I)^{-1} Y^{(t)} R^\top$$

Solving Alternative Least Square

Similarly, we can get

$$Y^{(t+1)} = (X^{(t+1)} {X^{(t+1)}}^\top + \lambda I)^{-1} X^{(t+1)} R$$

Therefore, the iterative scheme can be posed as follows:

1. $X^{(t+1)} = \left(Y^{(t)} {Y^{(t)}}^\top + \lambda I \right)^{-1} Y^{(t)} R^\top$
2. $Y^{(t+1)} = \left(X^{(t+1)} {X^{(t+1)}}^\top + \lambda I \right)^{-1} X^{(t+1)} R$

Matrix Factorization

Let $\Omega = \{(i,j) \mid R_{i,j} > 0\}$ and

$$[\mathcal{P}_\Omega(A)]_{i,j} = \begin{cases} A_{i,j} & \text{if } (i,j) \in \Omega \\ 0 & \text{if } (i,j) \notin \Omega \end{cases}$$

Problem

$$\min_{X,Y} \|\mathcal{P}_\Omega(R - X^\top Y)\|_F^2 + \lambda \left(\|X\|_F^2 + \|Y\|_F^2\right)$$

where

- $R \in \mathbb{R}^{m\times n}$, $X \in \mathbb{R}^{k\times m}$, and $Y \in \mathbb{R}^{k\times n}$
- $\lambda > 0$ is a parameter

Matrix Factorization

Problem

$$\min_{X,Y} \|\mathcal{P}_{\Omega}(R - X^{\top}Y)\|_F^2 + \lambda \left(\|X\|_F^2 + \|Y\|_F^2\right)$$

Its equivalence to

$$\min_{x_i,y_j} \sum_{(i,j)\in\Omega} (r_{i,j} - x_i^{\top} y_j)^2 + \lambda \left(\sum_{i=1}^{m} \|x_i\|_2^2 + \sum_{i=1}^{n} \|y_j\|_2^2 \right)$$

where $x_i, y_j \in \mathbb{R}^{k\times 1}$.

Solving Alternative Least Square

For fixed i, we define

$$F(x_i) = \sum_{(i,j)\in\Omega} (r_{i,j} - y_j^\top x_i)^2 + \sum_{i=1}^{m} \|x_i\|_2^2$$

Then

$$\nabla F(x_i) = -2 \sum_{(i,j)\in\Omega} y_j(r_{i,j} - y_j^\top x_i) + 2\lambda x_i$$

Letting $\nabla F(x_i) = 0$, we have

$$\left(\sum_{(i,j)\in\Omega} y_j y_j^\top + \lambda I \right) x_i = r_{i,j} y_j$$

Solving Alternative Least Square

Algorithm (ALS)

1. Initial $x_1, x_2, ..., x_m, y_1, y_2, ..., y_n$.
2. For $i = 1, 2, ..., m$:

$$\left(\sum_{(i,j)\in\Omega} y_j y_j^\top + \lambda I \right) x_i = r_{i,j} y_j$$

3. For $j = 1, 2, ..., n$:

$$\left(\sum_{(i,j)\in\Omega} x_i x_i^\top + \lambda I \right) y_j = r_{i,j} x_i$$

4. Repeat 2, 3 until convergence.

Stochastic Gradient Descent (SGD)

Algorithm (SGD)

1. Initial: $x_i, y_j \in \mathbb{R}^{k\times 1}$ for $i = 1, 2, ..., n, j = 1, 2, ..., m$
2. Randomly pick $(i, j) \in \Omega$
3. Calculate residual with regularization term:
$$\tilde{r}_{i,j} = r_{i,j} - x_i^\top y_j + \lambda$$
4. SGD update:
$$x_i \leftarrow x_i + \rho \tilde{r}_{i,j} x_i$$
$$y_j \leftarrow y_j + \rho \tilde{r}_{i,j} y_j$$
where $\rho > 0$ is learning rate.
5. Repeat 2-4 until convergence.

Soft Impute Alternative Least Square (SIALS)

Problem

$$\min_{X,Y} \|\mathcal{P}_\Omega(R - X^\top Y)\|_F^2 + \lambda \left(\|X\|_F^2 + \|Y\|_F^2\right)$$

$$\mathcal{P}_\Omega(R - X^\top Y) = \mathcal{P}_\Omega(R) - \mathcal{P}_\Omega(X^\top Y) + X^\top Y - X^\top Y$$

Let

$$\tilde{R} = \mathcal{P}_\Omega(R) - \mathcal{P}_\Omega(X^\top Y) + X^\top Y$$

Then the problem becomes

$$\min_{X,Y} \|\tilde{R} - X^\top Y\|_F^2 + \lambda \left(\|X\|_F^2 + \|Y\|_F^2\right)$$

Soft Impute Alternative Least Square (SIALS)

SIALS method

Given an initial $X^{(0)}, Y^{(0)}$, for $t = 0, 1, ...$, we solve the following two sub-problem alternatingly:

$$\begin{aligned}
\tilde{R}^{(t+\frac{1}{2})} &= \mathcal{P}_\Omega(R) - \mathcal{P}_\Omega({X^{(t)}}^\top Y^{(t)}) + {X^{(t)}}^\top Y^{(t)} \\
X^{(t+1)} &= \left(Y^{(t)} {Y^{(t)}}^\top + \lambda I\right)^{-1} Y^{(t)} \tilde{R}^\top \\
\tilde{R}^{(t+1)} &= \mathcal{P}_\Omega(R) - \mathcal{P}_\Omega({X^{(t+1)}}^\top Y^{(t)}) + {X^{(t+1)}}^\top Y^{(t)} \\
Y^{(t+1)} &= \left(X^{(t+1)} {X^{(t+1)}}^\top + \lambda I\right)^{-1} X^{(t+1)} \tilde{R}
\end{aligned}$$

We iterate until convergence is achieved.

Solving SIALS by SVD

Idea (rank k approximation) [1]

$$R = U\Sigma V^\top \approx U_k D_k^2 V_k = (D_k U_k^\top)^\top (D_k V_k^\top) = X^\top Y$$

Goal

Given $U^{(0)} \in \mathbb{R}^{m\times k}$ with orthonormal columns, $D^{(0)} = I_k$, and $V^{(0)} = O_{n\times k}$. We use the Iterative method to find the suitable $X = D^{(t)} {U^{(t)}}^\top$, $Y = D^{(t)} {V^{(t)}}^\top$ such that

$$\min_{X,Y} \|P_\Omega(R - X^\top Y)\|_F^2 + \lambda \left(\|X\|_F^2 + \|Y\|_F^2\right)$$

[1] T. Hastie, R. Mazumder, J.-D. Lee and R. Zadeh, Matrix Completion and Low-Rank SVD via Fast Alternating Least Squares, Journal of Machine Learning Research (2015).

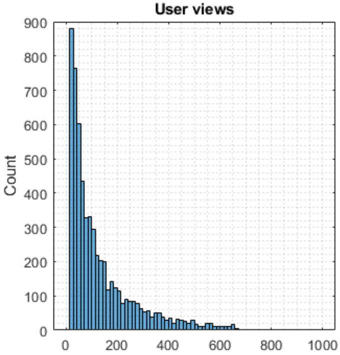
Solving SIALS by SVD

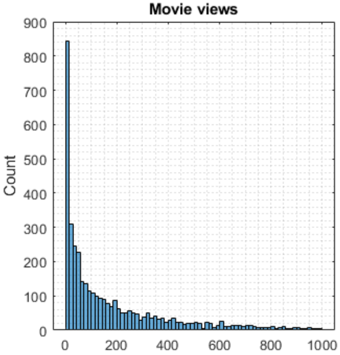
Algorithm (SIALS)

1. Initial: $R \in \mathbb{R}^{m\times n}, U \in \mathbb{R}^{m\times k}$ with orthonormal columns, $D = I$, $V = O_{n\times k}$, $X = DU^\top$ and $Y = DV^\top$.
2. $S \leftarrow \mathcal{P}_\Omega(R) - \mathcal{P}_\Omega(X^\top Y)$
3. $\tilde{X} \leftarrow \left(D^2 + \lambda I\right)^{-1} DV^\top S^\top + \left(D^2 + \lambda I\right)^{-1} D^2 X$
4. $U, \tilde{D} \leftarrow \mathrm{SVD}(D\tilde{X})$, $X \leftarrow \tilde{D}U^\top$
5. $S \leftarrow \mathcal{P}_\Omega(R) - \mathcal{P}_\Omega(X^\top Y)$
6. $\tilde{Y} \leftarrow \left(D^2 + \lambda I\right)^{-1} DU^\top S + \left(D^2 + \lambda I\right)^{-1} D^2 Y$
7. $D, V \leftarrow \mathrm{SVD}(\tilde{D}\tilde{Y})$, $Y \leftarrow \tilde{D}V^\top$
8. Repeat 2-7 until convergence.
9. Output: U, V, D

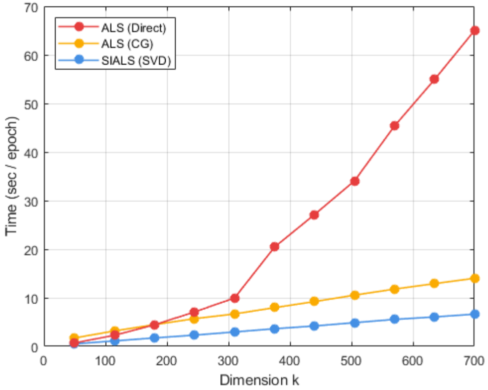
Information of Dataset: ml-1m

- Number of data: 900,188
- Number of user: 6,040
- Number of movie: 3,952
- Sparsity: 3.77%





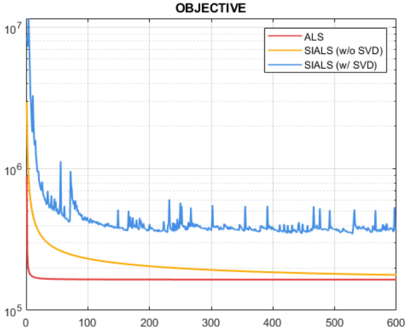
Time Elapsed



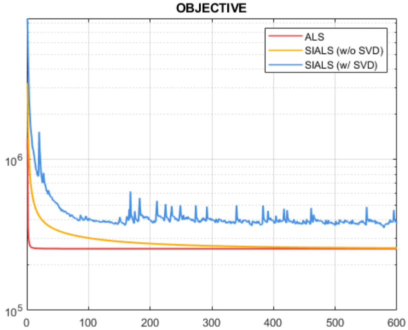
Remark

Direct method: $O(|\Omega|k^2 + (m+n)k^3)$

Objective



$\lambda = 5,\ d = 50$



$\lambda = 20,\ d = 50$

Metrics

We use the following metrics to evaluate our methods:

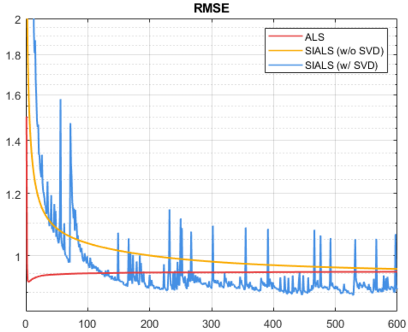
1. **Root Mean Square Error (RMSE):**

$$\mathrm{RMSE}_{\Omega}(X, Y) = \sqrt{\frac{\sum_{(i,j)\in\Omega} |X_{i,j} - Y_{i,j}|^2}{|\Omega|}}$$

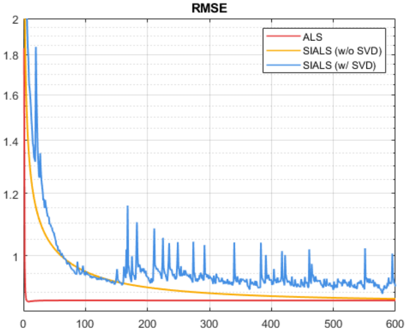
2. **Mean Absolute Error (MAE)**:

$$\mathrm{MAE}_{\Omega}(X, Y) = \frac{\sum_{(i,j)\in\Omega} |X_{i,j} - Y_{i,j}|}{|\Omega|}$$

Evaluation



$\lambda = 5,\ d = 50$



$\lambda = 20,\ d = 50$

Evaluation

Metrics	RMSE		MAE	
Type	Training	Testing	Training	Testing
ALS	0.5542	0.9545	0.4363	0.7385
SIALS (w/o SVD)	0.5745	0.9624	0.4517	0.7441
SIALS (w/ SVD)	0.8928	**0.9099**	0.7017	0.7143

$\lambda = 5$, $d = 50$ with 600 epochs

Metrics	RMSE		MAE	
Type	Training	Testing	Training	Testing
ALS	0.7288	**0.8778**	0.5727	0.6889
SIALS (w/o SVD)	0.7330	0.8820	0.5749	0.6913
SIALS (w/ SVD)	0.8980	0.9147	0.7101	0.7229

$\lambda = 20$, $d = 50$ with 600 epochs

ml-10m dataset

- Number of data: 9,000,048
- Number of user: 71,567
- Number of movie: 65,133
- Sparsity: 0.19%

Metrics	RMSE		MAE	
Type	Training	Testing	Training	Testing
ALS	0.7287	**0.8453**	0.5453	0.6479
SIALS (w/ SVD)	0.9301	0.9454	0.7171	0.7286

$\lambda = 50$, $d = 100$ with 500 epochs

Prediction

We compute $\hat{r}_{i,j} = x_i^\top y_j$, and list the prediction as the following:

(User, Movie)	Rating	Prediction
(308, 1707)	2	1.5211
(990, 89)	3	2.7883
(2247, 2291)	3	4.0016
(2454, 595)	3	3.8591
(2853, 3363)	4	3.7680
(3067, 703)	1	1.0212
(3317, 3793)	4	4.1885
(3727, 2259)	2	2.2826
(4796, 2761)	4	4.1454
(5451, 969)	5	4.4651

Conclusion

In this project, we implement ALS, SIALS to do the matrix factorization. We use the direct method and CG to solve ALS and use SVD to solve SIALS and then apply it to the recommender system.

Reference

1. Trevor Hastie, Rahul Mazumder, Jason D. Lee and Reza Zadeh, Matrix Completion and Low-Rank SVD via Fast Alternating Least Squares, Journal of Machine Learning Research 16 (2015), 3367-3402.
2. Yehuda Koren, Robert Bell and Chris Volinsky, Matrix Factorization Techniques for Recommender Systems, *IEEE* Computer Society 42 (2009), 30-37.

THE END

Thanks for listening!