1. Prepare the data by filling unknown ratings with zero

$$Y = \begin{bmatrix} 4 & 0 \\ 0 & 0 \\ 3 & 5 \end{bmatrix} \qquad Y^T = \begin{bmatrix} 4 & 0 & 3 \\ 0 & 0 & 5 \end{bmatrix}$$

User-item Item-user

2. Perform clustering on users and items

$\{user\_group1, user\_gruop2\}$ $\{Item\_group1, Item\_gruop2\}$

- User 1
- User 2

- Item 3

- Item 1

- Item 2

This stage still has room for improvement, such as the clustering algorithm and auto-clustering methods.

3. Combine the user clusters and item clusters into a "favorite transaction" set

User 1: $\{Item1\}$ $\{User\_group2, Item1, Item\_group1\}$

User 2: $\{\ \}$ Cold start user $\{User\_group1\}$

User 2: $\{Item2\}$ $\{User\_group1, Item2, Item\_group2\}$

This is how we solve the cold-start problem; when a cold-start user has a segment, we can create an item signal or a group of item signals for them.

A "favorite transaction" contains items with a score greater than or equal to 80%.

4. Mine for rules from the "favorite transaction" set using the FP-growth algorithm

1. $\{\text{User\_group2}\} \rightarrow \{\text{Item\_group2}\}$ Confidence = 0.5
2. $\{\text{User\_group2, item1}\} \rightarrow \{\text{Item2}\}$ Confidence = 0.3
3. $\{\text{Item2}\} \rightarrow \{\text{Item1}\}$ Confidence = 0.6
4. $\{\text{User\_group1}\} \rightarrow \{\text{Item1}\}$ Confidence = 0.7

There are two popular association rule mining algorithms, Apriori and FP-growth, but Apriori is computationally expensive, so we use FP-growth instead.

5. Map the resulting rules to each user

User 1: $\{\text{User\_group2, Item1, Item\_group1}\}$ → Rule: 1, 2

User 2: $\{\text{User\_group1}\}$ Cold start user → Rule: 4

User 3: $\{\text{User\_group1, Item2, Item\_group2}\}$ → Rule: 1,3, 4

6. Create an Item-Signal and a Group-of-Item-Signal Matrix for each user

$$M = \begin{bmatrix} 0 & 0.3 \\ 0 & 0.6 \\ 0.6 + 0.7 & 0 \end{bmatrix} \quad N = \begin{bmatrix} 0 & 0.5 \\ 0 & 0 \\ 0 & 0.5 \end{bmatrix}$$

If the number of signals is greater than one, they are summed together.

7. Normalize the signals to a range of [0,1] to prevent the gradient from exploding

$$\mathrm{M} = \begin{bmatrix} 0 & 0.3 \\ 0 & 0.6 \\ 0.6 + 0.7 & 0 \end{bmatrix} \quad \mathrm{N} = \begin{bmatrix} 0 & 0.5 \\ 0 & 0 \\ 0 & 0.5 \end{bmatrix} \qquad \mathrm{M\_norm} = \begin{bmatrix} 0 & 0.23 \\ 0 & 0.46 \\ 1 & 0 \end{bmatrix} \quad \mathrm{N\_norm} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

8. Formulate a new loss function

$$\sum (r_{ui} - (\mu + b_i + b_u + p_u^T q_i + \boldsymbol{\alpha_u \cdot m_{ui} + \beta_u \cdot n_{ui}}))^2 - \lambda \sum Reg(\theta)$$

9. Solve for all parameters using gradient descent.

$$\begin{aligned} q_i &\leftarrow q_i + \eta(e_{ui} \cdot p_u - \lambda q_i) & \mu &\leftarrow \mu + \eta(e_{ui}) & \alpha_u &\leftarrow \alpha_u + \eta(e_{ui} \cdot m_u - \lambda \alpha_u) \\ p_u &\leftarrow p_u + \eta(e_{ui} \cdot q_i - \lambda p_u) & b_i &\leftarrow b_i + \eta(e_{ui} - \lambda b_i) & \beta_u &\leftarrow \beta_u + \eta(e_{ui} \cdot n_u - \lambda \beta_u) \\ & & b_u &\leftarrow b_u + \eta(e_{ui} - \lambda b_u) & & \end{aligned}$$

10. Predict the final prediction.

$$\hat{r}_{ui} = \mu + b_i + b_u + p_u^T q_i + \boldsymbol{\alpha_u \cdot m_{ui} + \beta_u \cdot n_{ui}}$$

Movielen-small: Epochs=100, Learning rate=0.001, lambda=0.001,K=50,batch_size=512

Solver: ADAM

	Dataset	Model	K	LR	Lambda	Epochs	MAP@5_mean	MAP@5_std	MAP@5_cold_mean	MAP@5_cold_std	MAP@5_not_cold_mean	MAP@5_not_cold_std
0	ml_small	CAR_MF	50	0.001	0.001	100	0.104102	0.033693	0.074608	0.055472	0.109319	0.032127
1	ml_small	CAR_MF+Bias	50	0.001	0.001	100	0.175880	0.048358	0.188845	0.057908	0.174010	0.057192
2	ml_small	MF	50	0.001	0.001	100	0.041455	0.018774	0.046774	0.020444	0.040803	0.023736
3	ml_small	MF+Bias	50	0.001	0.001	100	0.078060	0.074426	0.110321	0.075710	0.068660	0.073636