Factorization-Based Data Modeling
Practical Work 3

Xiaoxuan HEI, Yuru HE

February 2020

1 Coupled factorization model

This dataset has a main tensor X_1 of size $164 \times 168 \times 5$, which encapsulates user-location-activity informations, where $X_1(i; j; k) = 1$ if the user i visits location j and performs activity k there and $X_1(i; j; k) = 0$ otherwise. The dataset also includes additional side information: the user-location preferences matrix X_2, the location-feature matrix X_3, the user-user similarity matrix X_4, and the activity-activity matrix X_5. Our aim is to predict the missing parts of X_1.

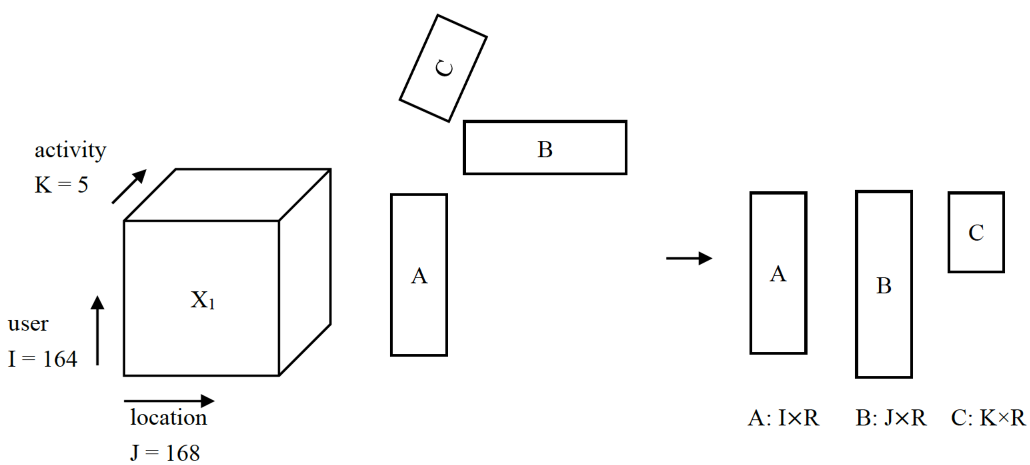

Figure 1: Factorization of X_1

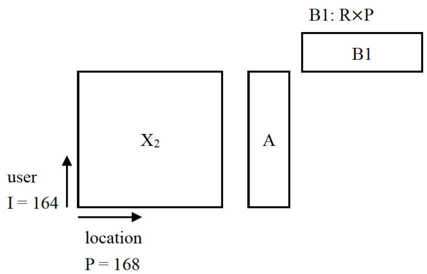

Figure 2: Factorization of X_2

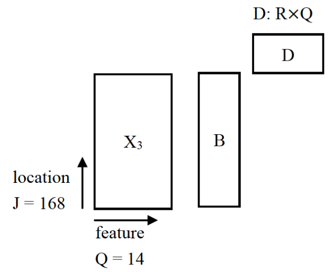

Figure 3: Factorization of X_3

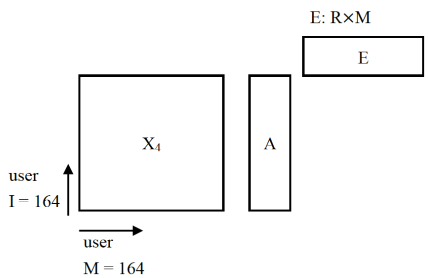

Figure 4: Factorization of X_4

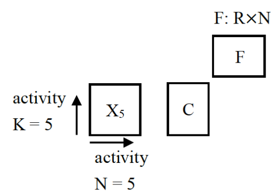

Figure 5: Factorization of X_5

2 Write down the cost function

$$
\begin{aligned}
F =&\lambda_1 \sum_{i=1}^{164}\sum_{j=1}^{168}\sum_{k=1}^{5} d_{\beta_1}\left(x_{1,ijk} \parallel \hat{x}_{1,ijk}\right) + \lambda_2 \sum_{i=1}^{164}\sum_{p=1}^{168} d_{\beta_2}\left(x_{2,ip} \parallel \hat{x}_{2,ip}\right) + \lambda_3 \sum_{j=1}^{168}\sum_{q=1}^{14} d_{\beta_3}\left(x_{3,jq} \parallel \hat{x}_{3,jq}\right) + \\
&\lambda_4 \sum_{i=1}^{164}\sum_{m=1}^{164} d_{\beta_4}\left(x_{4,im} \parallel \hat{x}_{4,im}\right) + \lambda_5 \sum_{k=1}^{5}\sum_{n=1}^{5} d_{\beta_5}\left(x_{5,kn} \parallel \hat{x}_{5,kn}\right) \\
=&\lambda_1 \sum_{i=1}^{164}\sum_{j=1}^{168}\sum_{k=1}^{5} \left(\frac{x_{1,ijk}^{\beta_1}}{\beta_1} - \frac{x_{1,ijk}\hat{x}_{1,ijk}^{\beta_1-1}}{\beta_1-1} + \frac{\hat{x}_{1,ijk}^{\beta_1}}{\beta_1}\right) + \lambda_2 \sum_{i=1}^{164}\sum_{p=1}^{168} \left(\frac{x_{2,ip}^{\beta_2}}{\beta_2} - \frac{x_{2,ip}\hat{x}_{2,ip}^{\beta_2-1}}{\beta_2-1} + \frac{\hat{x}_{2,ip}^{\beta_2}}{\beta_2}\right) + \\
&\lambda_3 \sum_{j=1}^{168}\sum_{q=1}^{14} \left(\frac{x_{3,jq}^{\beta_3}}{\beta_3} - \frac{x_{3,jq}\hat{x}_{3,jq}^{\beta_3-1}}{\beta_3-1} + \frac{\hat{x}_{3,jq}^{\beta_3}}{\beta_3}\right) + \lambda_4 \sum_{i=1}^{164}\sum_{m=1}^{164} \left(\frac{x_{4,im}^{\beta_4}}{\beta_4} - \frac{x_{4,im}\hat{x}_{4,im}^{\beta_4-1}}{\beta_4-1} + \frac{\hat{x}_{4,im}^{\beta_4}}{\beta_4}\right) + \\
&\lambda_5 \sum_{k=1}^{5}\sum_{n=1}^{5} \left(\frac{x_{5,kn}^{\beta_5}}{\beta_5} - \frac{x_{5,kn}\hat{x}_{5,kn}^{\beta_5-1}}{\beta_5-1} + \frac{\hat{x}_{5,kn}^{\beta_5}}{\beta_5}\right)
\end{aligned}
$$

3 Explain why our model makes sense

To fill in entries in the tensor X_1, we propose a PARAFAC-style tensor decomposition framework to integrate the tensor with the additional matrices for a regularized decomposition. First, we decomposeX_1 for some low-dimensional representation w.r.t. each tensor entity (i.e. users A, locations B and activities C). After such low-dimensional representations are obtained, we can reconstruct the tensor by filling all the missing entries in the tensor. X_2 shares the the low-dimensional users representation A. X_3 shares the the low-dimensional location representation B. X_4 shares the the low-dimensional users representation A. X_5 shares the the low-dimensional activities representation C. According to the shared and complete representations, we can fill in X_1.

4 Develop the multiplicative update rules algorithm

$$\frac{\partial F}{\partial a_{ir}} = \lambda_1 \sum_{j=1}^{168}\sum_{k=1}^{5}(\hat{x}_{1,ijk}^{\beta_1-1} - x_{1,ijk}\hat{x}_{1,ijk}^{\beta_1-2})b_{jr}c_{kr} + \lambda_2 \sum_{p=1}^{168}(\hat{x}_{2,ip}^{\beta_2-1} - x_{2,ip}\hat{x}_{2,ip}^{\beta_2-2})b1_{rp} + \lambda_4 \sum_{m=1}^{164}(\hat{x}_{4,im}^{\beta_4-1} - x_{4,im}\hat{x}_{4,im}^{\beta_4-2})e_{rm}$$

$$\frac{\partial F}{\partial b_{jr}} = \lambda_1 \sum_{i=1}^{164}\sum_{k=1}^{5}(\hat{x}_{1,ijk}^{\beta_1-1} - x_{1,ijk}\hat{x}_{1,ijk}^{\beta_1-2})a_{ir}c_{kr} + \lambda_3 \sum_{q=1}^{168}(\hat{x}_{3,jq}^{\beta_3-1} - x_{3,jq}\hat{x}_{3,jq}^{\beta_3-2})d_{rq}$$

$$\frac{\partial F}{\partial c_{kr}} = \lambda_1 \sum_{i=1}^{164}\sum_{j=1}^{168}(\hat{x}_{1,ijk}^{\beta_1-1} - x_{1,ijk}\hat{x}_{1,ijk}^{\beta_1-2})a_{ir}b_{jr} + \lambda_5 \sum_{n=1}^{5}(\hat{x}_{5,kn}^{\beta_5-1} - x_{5,kn}\hat{x}_{5,kn}^{\beta_5-2})f_{rn}$$

$$\frac{\partial F}{\partial b1_{rp}} = \lambda_2 \sum_{i=1}^{164}(\hat{x}_{2,ip}^{\beta_2-1} - x_{2,ip}\hat{x}_{2,ip}^{\beta_2-2})a_{ir} \qquad \frac{\partial F}{\partial d_{rq}} = \lambda_3 \sum_{j=1}^{168}(\hat{x}_{3,jq}^{\beta_3-1} - x_{3,jq}\hat{x}_{3,jq}^{\beta_3-2})b_{jr}$$

$$\frac{\partial F}{\partial e_{rm}} = \lambda_4 \sum_{i=1}^{164}(\hat{x}_{4,im}^{\beta_4-1} - x_{4,im}\hat{x}_{4,im}^{\beta_4-2})a_{ir} \qquad \frac{\partial F}{\partial f_{rn}} = \lambda_5 \sum_{k=1}^{5}(\hat{x}_{5,kn}^{\beta_5-1} - x_{5,kn}\hat{x}_{5,kn}^{\beta_5-2})c_{kr}$$

After getting the Partial derivative, we can get the update formula according to:

$$w_{ir}^{(t)} = w_{ir}^{(t-1)} - \eta\frac{\partial f}{\partial w_{ir}}$$

Let's take e_{rm} for example,

$$e_{rm}^{(t)} = e_{rm}^{(t-1)} - \eta\lambda_4 \sum_{i=1}^{164}(\hat{x}_{4,im}^{\beta_4-1} - x_{4,im}\hat{x}_{4,im}^{\beta_4-2})a_{ir}$$

$$\eta = \frac{e_{rm}^{(t-1)}}{\lambda_4 \sum_{i=1}^{164}(\hat{x}_{4,im}^{\beta_4-1}a_{ir})}$$

$$e_{rm}^{(t)} = \frac{e_{rm}^{(t-1)} \sum_{i=1}^{164} x_{4,im} \hat{x}_{4,im}^{\beta_4-2} a_{ir}}{\lambda_4 \sum_{i=1}^{164} (\hat{x}_{4,im}^{\beta_4-1} a_{ir})}$$

The rest can be obtained in the same way.

5 Implement our algorithm in MATLAB

When $\beta_{1:5} = 1, \lambda_{2:5} = 1$,

$$A \leftarrow A \odot \frac{\sum_{i,j,k} x_{1,ijk}/a_{ir}}{\sum_{i,j,k} b_{jr} c_{kr}} + A \odot \frac{(X_2/\hat{X}_2) B1^T}{O_2 B1^T} + A \odot \frac{(X_4/\hat{X}_4) E^T}{O_4 E^T}$$

$$B \leftarrow B \odot \frac{\sum_{i,j,k} x_{1,ijk}/b_{jr}}{\sum_{i,j,k} a_{ir} c_{kr}} + A \odot \frac{(X_2/\hat{X}_2) D^T}{O_3 D^T}$$

$$C \leftarrow C \odot \frac{\sum_{i,j,k} x_{1,ijk}/c_{kr}}{\sum_{i,j,k} a_{ir} b_{jr}} + A \odot \frac{(X_5/\hat{X}_5) F^T}{O_5 F^T}$$

$$B1 \leftarrow B1 \odot \frac{A^T (X_2/\hat{X}_2)}{A^T O_2} \qquad D \leftarrow D \odot \frac{B^T (X_3/\hat{X}_3)}{B^T O_3}$$

$$E \leftarrow E \odot \frac{A^T (X_4/\hat{X}_4)}{A^T O_4} \qquad F \leftarrow F \odot \frac{C^T (X_5/\hat{X}_5)}{C^T O_5}$$

When $\beta_{1:5} = 2, \lambda_{2:5} = 1$,

$$A \leftarrow A \odot \frac{\sum_{i,j,k} x_{1,ijk} b_{jr} c_{kr}}{\sum_{i,j,k} \hat{x}_{1,ijk} b_{jr} c_{kr}} + A \odot \frac{X_2 B1^T}{\hat{X}_2 B1^T} + A \odot \frac{X_4 E^T}{\hat{X}_4 E^T}$$

$$B \leftarrow B \odot \frac{\sum_{i,j,k} x_{1,ijk} a_{ir} c_{kr}}{\sum_{i,j,k} \hat{x}_{1,ijk} a_{ir} c_{kr}} + B \odot \frac{X_3 D^T}{\hat{X}_3 D^T}$$

$$C \leftarrow C \odot \frac{\sum_{i,j,k} x_{1,ijk} a_{ir} b_{jr}}{\sum_{i,j,k} \hat{x}_{1,ijk} b_{jr} c_{kr}} + C \odot \frac{X_5 F^T}{\hat{X}_5 F^T}$$

$$B1 \leftarrow B1 \odot \frac{A^T X_2}{A^T \hat{X}_2} \qquad D \leftarrow D \odot \frac{B^T X_3}{B^T \hat{X}_3}$$

$$E \leftarrow E \odot \frac{A^T X_4}{A^T \hat{X}_4} \qquad F \leftarrow F \odot \frac{C^T X_5}{C^T \hat{X}_5}$$

where $\odot$ denotes element-wise multiplication and $\div$ denotes element-wise division. Here, we also define the matrix $O_{2:5}$ of size $X_{2:5}$, such that each element of $O_{2:5}$ will be equal to 1.

Then we can get the update formula for different β in the same way.

As for λ, the more complete the data of the matrix is, the more weight we give.

6 Experiment results

The cost function doesn't converge, but when we give appropriate $\lambda_{1:5}$, it can approach convergence. .In our dataset, matrix X_3, X_4 and X_5 are more complete, so we give them bigger λ. X_1 and X_2 are very sparse, we give them smaller λ.

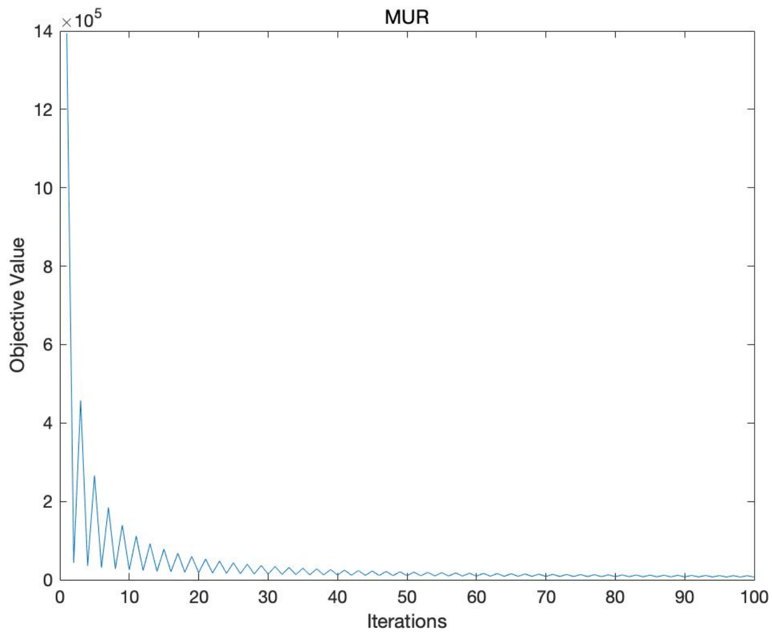

Figure 6: Cost function when $\beta_{1:5} = 1$

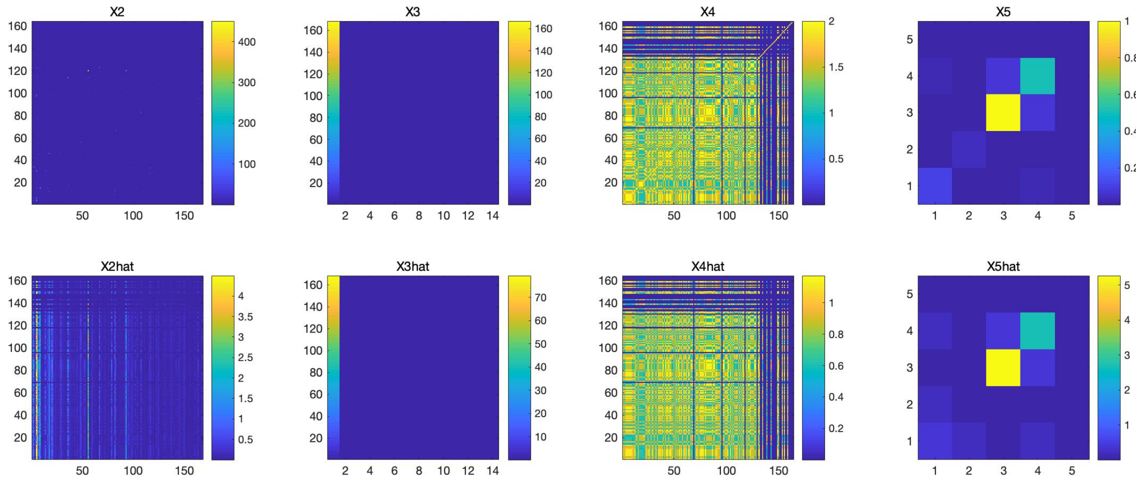

Figure 7: The similarity between $X_{2:5}$ and $\hat{X}_{2:5}$ when $\beta_{1:5} = 1$

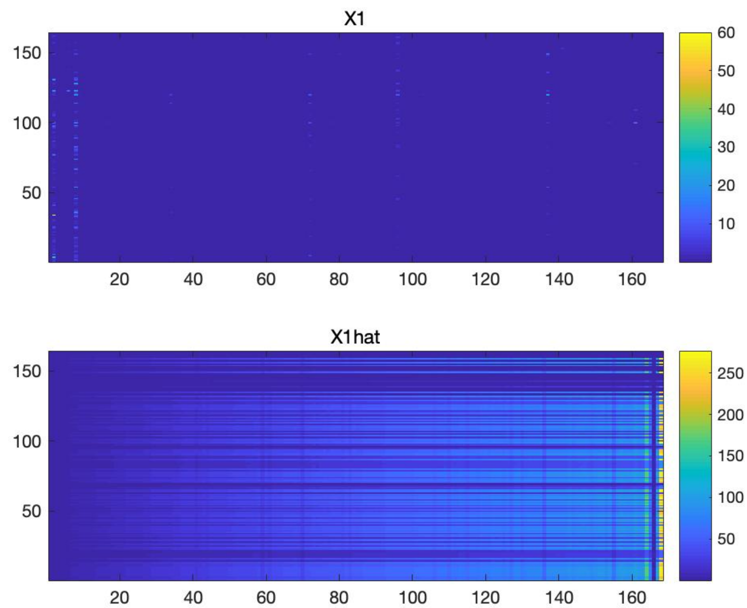

Figure 8: A part of X_1 when $\beta_{1:5} = 1$

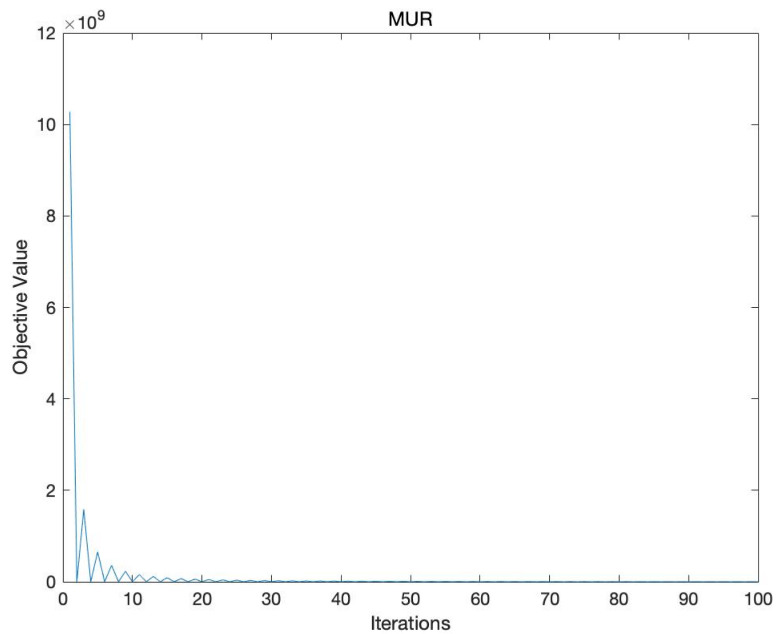

Figure 9: Cost function when $\beta_{1:5} = 2$

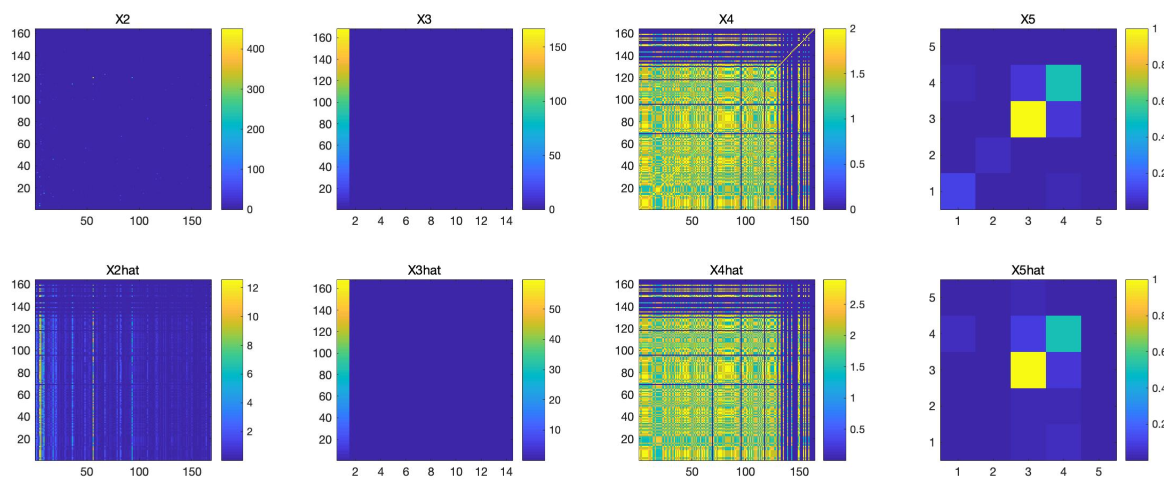

Figure 10: The similarity between $X_{2:5}$ and $\hat{X}_{2:5}$ when $\beta_{1:5} = 2$

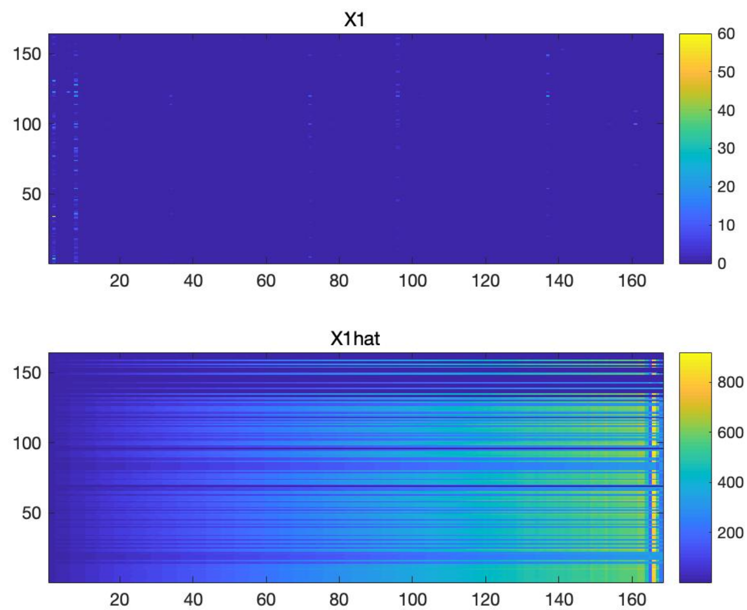

Figure 11: A part of X_1 when $\beta_{1:5} = 2$