Unsupervised/Semi-Supervised Binary Tensor Factorization on dnations data

Zhuoyan Xu

1 Unsupervised Tensor factorization

The general model is:

$$logit\left\{\mathbb{E}\left[\mathcal{X}^{d_1 d_2 d_3}\right]\right\} = \mathcal{G}^{r_1 r_2 r_3} \times_1 A^{d_1 r_1} \times_2 B^{d_2 r_2} \times_3 C^{d_3 r_3}$$

Where d_1, d_2, d_3 is dimension of tensor. The r_1, r_2, r_3 is the dimension of low rank tensor.

For each of the factor matrix and core tensor, we can unfold or vectorize it to deduce a matrix form product, then use GLM to compute.

1.1 Matrix form GLM

Before we deduce, we first introduce a matrix form GLM. Consider

$$logit[E(Y))_{n*p}] = U_{n*p} = X_{n*R} \times \beta_{R*p}$$

Where $U_{n*p} = (u_1, \ldots, u_p)$, $\beta_{R*p} = (\beta_1, \ldots, \beta_p)$

We have:

$$\begin{aligned} u_1^{N*1} &= X^{n*R} \times \beta_1^{R*1} \\ u_2^{N*1} &= X^{n*R} \times \beta_2^{R*1} \\ &\vdots \\ u_p^{N*1} &= X^{n*R} \times \beta_P^{R*1} \end{aligned}$$

Then we implement GLM.

1.2 Vectorization

Then we define a vectorization of the tensor in this scenario: When we say vectorize a tensor G (3 dimension), we extract all its mode-3 fiber and then list it according first mode-2 then according to mode-1. In other words, when we count the vector elements, the element's third mode index change first, then second mode index, finally first mode index.

Here is a small example for understanding:

$$Vec(G^{2*2*2}) = \begin{bmatrix} G_{111} \\ G_{211} \\ G_{121} \\ G_{221} \\ G_{112} \\ G_{222} \\ G_{112} \\ G_{222} \end{bmatrix}$$

The reason why we define like this is this is how as.vector() in r works.
In R: Vec(G) = as.vector(G).

1.3 Optimization/Updating method

Consider the general model, we can deduce a matrix form equation, then apply a matrix form GLM.

In this section use Y to denote the logits $logit(\mathbb{E}(Y))$.

1.3.1 The update of factor matrix A,B,C

$$Y^{d_1 d_1 d_3} = G^{r_1 r_2 r_3} \times_1 A^{d_1 r_1} \times_2 B^{d_2 r_2} \times_3 C^{d_3 r_3} = G_{AB}^{r_1 r_2 d_3} \times_3 C^{d_3 r_3}$$

Update A,B,C through:

$$Y_{(3)}^T = G_{AB(3)}^T C^T$$

Since for distinct modes in a series of multiplications, the order of the multiplication is irrelevant. Thus, A,B,C the same update way.

1.3.2 The update of core tensor $\mathcal{G}$

Consider

$$Y^{d_1 * d_2 * d_3} = G^{r_1 * r_2 * r_3} \times_1 A^{d_1 * r_1} \times_2 B^{d_2 * r_2} \times_3 C^{d_3 * r_3}$$

We have:

$$Y_{ijk} = \sum_{k'=1}^{r_3} \sum_{j'=1}^{r_2} \sum_{i'=1}^{r_1} G_{i'j'k'} A_{ii'} B_{jj'} C_{kk'}$$

Let

$$[M^{ijk}]_{i'j'k'} = A_{ii'} B_{jj'} C_{kk'}$$

$i = 1, \ldots, d_1; j = 1, \ldots, d_2; k = 1, \ldots, d_3$ There are $d_1 * d_2 * d_3$ M^{ijk} tensors totally.The dimension of M^{ijk} is $r_1 * r_2 * r_3$.

Then we have:

$$Y_{ijk} = \sum_{k'=1}^{r_3} \sum_{j'=1}^{r_2} \sum_{i'=1}^{r_1} G_{i'j'k'} M_{i'j'k'}^{ijk}$$

Since we have:

$$\text{Vec}\left(M^{ijk}\right) = C_{k:} \otimes B_{j:} \otimes A_{i:}$$

Thus:

$$\begin{aligned} Y_{ijk} &= \sum_{k'=1}^{r_3}\sum_{j'=1}^{r_2}\sum_{i'=1}^{r_1} G_{i'j'k'} M^{ijk}_{i'jk'} \\ &= Vec(G)^T \, \text{Vec}\left(M^{ijk}\right) = \text{Vec}\left(M^{ijk}\right)^T_{1*r_1r_2r_3} Vec(G)_{r_1r_2r_3*1} \end{aligned}$$

First step:

$$Y_{:jk} = \begin{bmatrix} y_{1jk} \\ y_{2jk} \\ \vdots \\ y_{d_1jk} \end{bmatrix}_{d_1*1} = \begin{bmatrix} Vec(M^{1jk})^T \\ Vec(M^{2jk})^T \\ \vdots \\ Vec(M^{d_1jk})^T \end{bmatrix}_{d_1*r_1r_2r_3} \times Vec(G)_{r_1r_2r_3*1} = Vec(M_{d_1jk}) \times Vec(G)$$

Second step:

$$Y_{::k} = \begin{bmatrix} y_{:1k} \\ y_{:2k} \\ \vdots \\ y_{:d_2k} \end{bmatrix}_{d_1d_2*1} = \begin{bmatrix} Vec(M_{d_11k}) \\ Vec(M_{d_12k}) \\ \vdots \\ Vec(M_{d_1d_2k}) \end{bmatrix}_{d_1d_2*r_1r_2r_3} \times Vec(G)_{r_1r_2r_3*1} = Vec(M_{d_1d_2k}) \times Vec(G)$$

Final step:

$$Y_{:::} = \begin{bmatrix} y_{::1} \\ y_{::2} \\ \vdots \\ y_{::d_3} \end{bmatrix}_{d_1d_2d_3*1} = \begin{bmatrix} Vec(M_{d_1d_21}) \\ Vec(M_{d_1d_22}) \\ \vdots \\ Vec(M_{d_1d_2d_3}) \end{bmatrix}_{d_1d_2d_3*r_1r_2r_3} \times Vec(G)_{r_1r_2r_3*1} = M^{long}_{d_1d_2d_3*r_1r_2r_3} \times Vec(G)_{r_1r_2r_3*1}$$

Where $Y_{:::}$ is the Vec(Y) vectorization of Y we defined brfore.

2 Semi-supervised Tensor factorization with covariate

Consider we have an extra covariate matrix $Y^{d1\times p}$(accounting for features), which contains the information of countries. We want to connect the membership matrix(or factor matrix) A and B with the information in Y. We came up with two models.

2.1 Treat Y as predictors

The first one, we treat Y as predictors. To connect the information of $\mathcal{X}$ and Y, we use matrix product to connect Y and one of $\mathcal{X}$'s factor matrices.

The general form is:

$$
\begin{aligned}
logit\left\{\mathbb{E}\left[\mathcal{X}^{d_1 d_2 d_3}\right]\right\} &= \mathcal{G}^{r_1 r_2 r_3} \times_1 A^{d_1 r_1} \times_2 B^{d_2 r_2} \times_3 C^{d_3 r_3} \\
A^{d_1 r_1} &= Y^{d_1 p} W^{p r_1}
\end{aligned}
$$

Where $\mathcal{G}$ is the low rank core tensor of factorization. A, B, C is three factor matrices. W is the regression coefficient matrix for Y on A. Under this scenario, $111 = p > d1 = 14$.

The degree of freedom of this model is

$$
r_1 r_2 r_3 + p r_1 + d_2 r_2 + d_3 r_3
$$

The total sample size is

$$
d_1 d_2 d_3 + d_1 p
$$

Note that we need to choose $r_1 < rank(Y)$.

2.2 Another view: Bilinear Model

We can write down the model in another view, which helps to compute:

$$
\begin{aligned}
logit\left\{\mathbb{E}\left[\mathcal{X}^{d_1 d_2 d_3}\right]\right\} &= \mathcal{C}^{p d_2 d_3} \times_1 Y^{d_1 p} \\
\mathcal{C}^{p d_2 d_3} &= \mathcal{G}^{r_1 r_2 r_3} \times_1 W^{p r_1} \times_2 B^{d_2 r_2} \times_3 C^{d_3 r_3}
\end{aligned}
$$

2.3 Optimization

The optimization is the same as previous with unsupervised model. We first came up with initialization of $\mathcal{C}$ given $\mathcal{X}$ and Y using glm. Then we use tucker decomposition to get initial value of $\mathcal{G}, W, B, C$.

The general form can be written as:

$$
logit\left\{\mathbb{E}\left[\mathcal{X}^{d_1 d_2 d_3}\right]\right\} = \mathcal{G}^{r_1 r_2 r_3} \times_1 (YW)^{d_1 r_1} \times_2 B^{d_2 r_2} \times_3 C^{d_3 r_3}
$$

Once we get the initial value, we can use glm to update $\mathcal{G}, W, B, C$ as we did before. The update of $\mathcal{G}$, B and C are same as unsupervised factorization.

Update of W when fixed $\mathcal{G}, B, C$ Once we fixed $\mathcal{G}, B, C$, consider

$$
\begin{aligned}
\mathcal{U}^{d_1 d_2 d_3} =& logit\left\{\mathbb{E}\left[\mathcal{X}^{d_1 d_2 d_3}\right]\right\} = \mathcal{G}^{r_1 r_2 r_3} \times_1 (YW)^{d_1 r_1} \times_2 B^{d_2 r_2} \times_3 C^{d_3 r_3} \\
=& \mathcal{G}_{BC}^{r_1 d_2 d_3} \times_1 (YW)^{d_1 r_1}
\end{aligned}
$$

Use U,G to denote the unfolded matrix of $\mathcal{U}, \mathcal{G}$ along mode 1. Thus we have:

$$
U^{d_1 \times d_2 d_3} = Y^{d_1 p} W^{p r_1} G^{r_1 \times d_2 d_3}
$$

Use tensor notation for above matrix product, we have:

$$
U^{d_1 \times d_2 d_3} = W^{p r_1} \times_1 Y^{d_1 p} \times_2 t(G)^{d_2 d_3 \times r_1}
$$

Consider element-wise formula for either tensor notation or matrix notation, we have:

$$\begin{aligned} U_{ij} &= \sum_{i'=1}^{p} \sum_{j'=1}^{r_1} W_{i'j'} Y_{ii'} t(G)_{jj'} \\ &= \sum_{i'=1}^{p} \sum_{j'=1}^{r_1} W_{i'j'} Y_{ii'} G_{j'j} \end{aligned}$$

Let

$$[N^{ij}]_{i'j'} = Y_{ii'} G_{j'j} = Y_{ii'} t(G)_{jj'}$$

where $i = 1, \ldots, d_1; j = 1, \ldots, d_2 d_3$ There are $d_1 * d_2 d_3$ N^{ij} tensors totally.The dimension of N^{ijk} is $p \times r_1$.

Then we have:

$$U_{ij} = \sum_{i'=1}^{p} \sum_{j'=1}^{r_1} W_{i'j'} N_{i'j'}^{ij}$$

Since we have:

$$\text{Vec}\left(N^{ij}\right) = t(G)_{j:} \otimes Y_{i:}$$

Thus:

$$\begin{aligned} U_{ij} &= \sum_{j'=1}^{r_1} \sum_{i'=1}^{p} W_{i'j'} N_{i'j'}^{ij} \\ &= Vec(N^{ij})_{1 \times pr_1}^{T} \, \text{Vec}\,(W)^{pr_1 \times 1} \end{aligned}$$

First step:

$$U_{:j} = \begin{bmatrix} U_{1j} \\ U_{2j} \\ \vdots \\ U_{d_1 j} \end{bmatrix}_{d_1 * 1} = \begin{bmatrix} Vec(N^{1j})^T \\ Vec(N^{2j})^T \\ \vdots \\ Vec(N^{d_1 j})^T \end{bmatrix}_{d_1 * pr_1} \times Vec(W)_{pr_1 * 1} = Vec(N_{d_1 j}) \times Vec(W)$$

Second step:

$$U_{::} = \begin{bmatrix} U_{:1} \\ U_{:2} \\ \vdots \\ U_{:d_2 d_3} \end{bmatrix}_{d_1 d_2 d_3 * 1} = \begin{bmatrix} Vec(N_{d_1 1}) \\ Vec(N_{d_1 2}) \\ \vdots \\ Vec(N_{d_1 d_2 d_3}) \end{bmatrix}_{d_1 d_2 d_3 * pr_1} \times Vec(W)_{pr_1 * 1} = N_{d_1 d_2 d_3 * pr_1}^{long} \times Vec(W)$$

The explict steps are shown below:

Algorithm 1 Semi-binary tensor decomposition

Input $\mathcal{X}, Y$, the shape of core tensor r_1, r_2, r_3, simulation times N
Use GLM to get $\mathcal{C}^{(1)}$, then use Tucker decomposition to get $W^{(1)}, B^{(1)}, C^{(1)}$
for $n = 1, 2, \ldots, N$ **do**
    (a) Update B,C use matrix form GLM, $B^{(n)} \leftarrow B^{(n+1)}, C^{(n)} \leftarrow C^{(n+1)}$
    (b) Update $\mathcal{G}$ using vectorization of tensor, $\mathcal{G}^{(n)} \leftarrow \mathcal{G}^{(n+1)}$
    (c) Update W using vectorization of matrix, $W^{(n)} \leftarrow W^{(n+1)}$
Output $\mathcal{G}, W, B, C$.

2.3.1 Simulation

Recalling the notarion form of bilinear model:

$$logit \left\{ \mathbb{E} \left[\mathcal{X}^{d_1 d_2 d_3} \right] \right\} = \mathcal{C}^{p d_2 d_3} \times_1 Y^{d_1 p}$$
$$\mathcal{C}^{p d_2 d_3} = \mathcal{G}^{r_1 r_2 r_3} \times_1 W^{p r_1} \times_2 B^{d_2 r_2} \times_3 C^{d_3 r_3}$$

I set

$$d_1 = d_2 = d_3 = 20$$
$$p = 5$$
$$r_1 = r_2 = r_3 = 2$$

To compare the estomation error of tensor $\mathcal{C}$ and tensor $\mathcal{U}$. I use the Frobenius norm of the difference of matrix to denote it. As we can see:

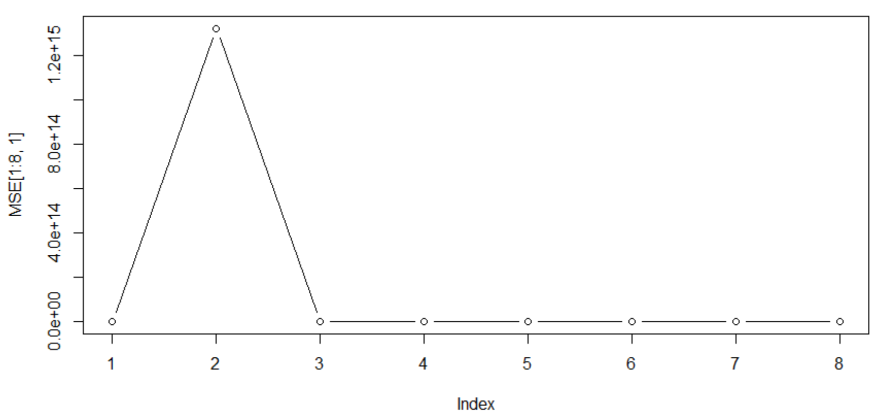

Figure 1: The error of C

As we can see, the error of second simulation is vary large(maybe because not convergence).

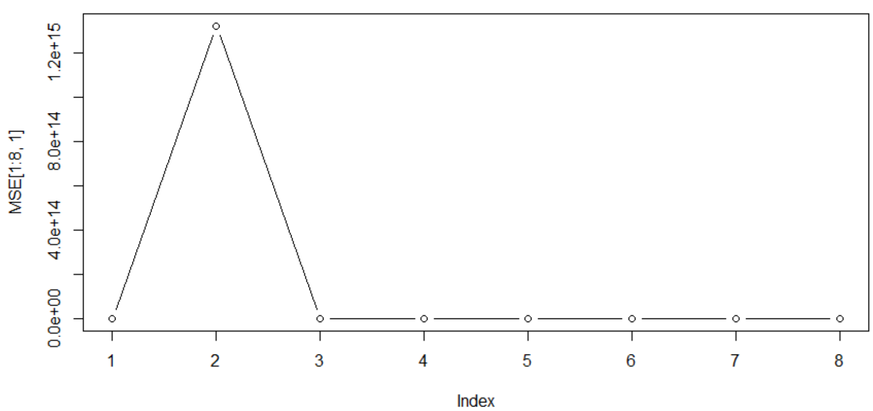

Figure 2: The error of U