Sparse clustering for multi-way array data

Miaoyan Wang, Feb 9, 2019

1 Theory

Definition 1 (Rank-$(r_1,\ldots,r_K)$ approximation)**.** Let $\mathcal{A}\in\mathbb{R}^{d_1\times\cdots\times d_K}$ be an order-K tensor and $r_1,\ldots,r_K$ be a set of integers satisfying $1\leq r_k\leq d_k$ for $k\in[K]$. The best rank-$(r_1,\ldots,r_K)$ approximation problem is to find a set of d_k-by-r_k matrices $\{\hat{U}^{(k)}\}$ with orthogonal columns and an $r_1\times r_2\times\cdots\times r_K$ core tensor $\hat{\mathcal{B}}$ such that

$$(\hat{\mathcal{B}},\hat{U}^{(k)})\in\mathop{\arg\min}_{\mathcal{B},U^{(k)}\in\mathcal{O}(d_k,r_k),k\in[K]}\left\|\mathcal{A}-\mathcal{B}\times_1\left(U^{(1)}\right)^T\times_2\cdots\times_K\left(U^{(K)}\right)^T\right\|_F.$$

It can be shown that the optimal $\hat{\mathcal{B}}$ is given by $\hat{\mathcal{B}}=\mathcal{A}\times_1\hat{U}^{(1)}\times_2\cdots\times_K\hat{U}^{(K)}$.

Definition 2 (Scaled Membership Matrix)**.** A matrix $U\in\mathbb{R}^{d\times r}$ is called a scaled membership matrix if the following two conditions hold:

1. All columns of U are orthonormal.
2. The elements of the i-th column of U are 0 or $\frac{1}{\sqrt{n_i}}$, where $n_i\in\mathbb{Z}^+$ for all $i\in[r]$ and $\sum_{i=1}^r n_i=d$.

We denote by $\mathcal{M}(d,r)$ the family of scaled membership matrices.

Theorem 1.1. *The optimization* (**??**) *subject to the constraints* $\{U^{(k)}\in\mathcal{M}(r,d):k=1,\ldots,K\}$ *is equivalent to the multi-way clustering problem with* $=0$.

2 Algorithm sketch

Algorithm for three-way sparse clustering:

Input: a tensor $\mathcal{A}\in\mathbb{R}^{d_1\times d_2\times d_3}$; the number of clusters (r_1,r_2,r_3) (i.e., there are r_1 clusters in the first mode, r_2 clusters in the second mode, and r_3 clusters in the third mode); sparse parameter λ.

Output: three membership matrices $\tilde{U}^{(1)}$, $\tilde{U}^{(2)}$, $\tilde{U}^{(3)}$.

1. **Initialize membership matrix.** Perform the k-mean clustering on the unfolded matrix $\mathcal{A}_{(1)}\in\mathbb{R}^{d_1\times(d_2d_3)}$. Store the resulting clustering of $\{1,\ldots,d_1\}$ into a *unscaled* membership matrix $\tilde{U}^{(1)}=[\![\tilde{U}^{(1)}(i,a)]\!]\in\{0,1\}^{d_1\times r_1}$; that is,

$$\tilde{U}^{(1)}(i,a)=\begin{cases}1 & \text{if index } i \text{ is in cluster } a,\\ 0 & \text{otherwise.}\end{cases}$$

In other words, $\tilde{U}^{(1)}$ is a dummy variable encoding a factor with r_1 levels (Recall what we have learned in STAT601).

Repeat the above procedure for $A_{(2)}$ and $A_{(3)}$ and record the resulting clustering into $\tilde{U}^{(2)}$ and $\tilde{U}^{(3)}$, respectively.

2. **Update cluster mean**.

$$\hat{\mathcal{B}} = \mathcal{A} \times_1 \tilde{U}^{(1)} \times_2 \tilde{U}^{(2)} \times_3 \tilde{U}^{(3)}.$$

   Note that we use the unscaled version of the membership matrix $\tilde{U}^{(1)}$, not the scaled membership matrix $\hat{U}^{(1)}$!

   Then perform **Soft-thresholding** on $\hat{\mathcal{B}}$.

$$\hat{\mathcal{B}}_{abc} \leftarrow \frac{\text{Soft-thresholding}(\hat{\mathcal{B}}_{abc}, \lambda)}{|\tilde{U}^{(1)}(:,a)||\tilde{U}^{(2)}(:,b)||\tilde{U}^{(3)}(:,c)|}, \quad \text{for all } (a,b,c) \in [r_1] \times [r_2] \times [r_3],$$

   where $|\tilde{U}^{(1)}(:,a)|$ denotes the sum of the i-th column in $\tilde{U}^{(1)}$ (or equivalently, the number of non-zero entries in the vector $\tilde{U}^{(1)}(:,a)$).

3. **Update group membership matrix**. Update $\tilde{U}^{(1)}$ using the following procedure:

   (Outer loop) for i from 1 to d_1:

   (Inner loop) for a from 1 to r_1:

   Compute $f(a) = \sum_{jk}(A_{ijk} - \hat{\mathcal{B}}_{abc})^2$ where b is the cluster index that j belongs to (based on the clustering on the second mode), c is the cluster index that k belongs to (based on the clustering on the third mode).

   (End of Inner loop) Find $a^* \in \{1, 2, \ldots, r_1\}$ which minimizes $f(a)$. Assign the index i to the cluster a^*.

   (End of Outer loop) Repeat the above procedure for all $i \in \{1, 2, \ldots, d_1\}$. Eventually, the output will be stored in a new membership matrix $\tilde{U}^{(1,\text{new})} \in \mathbb{R}^{d_1 \times r_1}$:

$$\tilde{U}^{(1,\text{new})}(i,a) = \begin{cases} 1 & \text{if the index } i \text{ is assigned to the cluster } a, \\ 0 & \text{otherwise.} \end{cases}$$

   Overwrite the previous membership matrix $\tilde{U}^{(1)}$ using this new matrix $\tilde{U}^{(1,\text{new})}$.

4. Repeat Step 2.
5. Repeat Step 3, but this time for $\tilde{U}^{(2)}$.
6. Repeat Step 2.
7. Repeat Step 3, but this time for $\tilde{U}^{(3)}$.

8. Repeat Steps 2-7 until convergence.

In each of the above 8 steps, record the objective function value

$$\text{Obj} = \frac{1}{2}\sum_{ijk}(\mathcal{A}_{ijk} - \hat{\mathcal{B}}_{abc})^2 + \lambda \sum_{abc} |\hat{\mathcal{B}}_{abc}|,$$

where in the first summand, a is the cluster that i belongs to (based on the clustering in the first mode), b is the cluster that j belongs to (based on the clustering in the second mode), c is the cluster that i belongs to (based on the clustering in the third mode).

Observe the change of Obj over iterations.

3 Preliminaries

Mode is a generalization of "row", "column", etc. For example, the first mode of a tensor means the "row", the second "mode" means the column, and the third "mode" means the 3rd direction.

The subscripts (1), (2), (3) specifies the **mode**.

$[n]$: for any integer n, I use the short-hand notation $[n]$ to represent the set $\{1, \ldots, n\}$.

i, j, k: index of the data entry.

a, b, c: index of the clusters.

$\mathcal{B}_{abc}$: the (a, b, c) entry in the tensor $\mathcal{B}$.

Both $\tilde{U}^{(1)}(i, a)$ and $\tilde{U}^{(1)}_{ia}$ represent the (i, a) entry in the matrix $\tilde{U}^{(1)}$.

Tensor-matrix multiplication Suppose $\mathcal{B} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$, $\mathcal{A} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, $U^{(1)} \in \mathbb{R}^{d_1 \times r_1}$, $U^{(2)} \in \mathbb{R}^{d_2 \times r_2}$,$U^{(3)} \in \mathbb{R}^{d_3 \times r_3}$. Then

$$\mathcal{B} = \mathcal{A} \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)}$$

means

$$\mathcal{B}_{abc} = \sum_{ijk} \mathcal{A}_{ijk} U^{(1)}_{ia} U^{(2)}_{jb} U^{(3)}_{kc}, \quad \text{for all } (a, b, c) \in [d_1] \times [d_2] \times [d_3].$$

Tensor unfolding. Check the R function unfold(..). Get a sense how the entries are re-arranged between $\mathcal{A}_{(1)}$, $\mathcal{A}_{(2)}$, $\mathcal{A}_{(3)}$ and $\mathcal{A}$.

Description of unfold(...):

Input: $\mathcal{A} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, mode

Output: $\mathcal{A}_{(1)} \in \mathbb{R}^{d_1 \times (d_2 d_3)}$ (if mode=1); $\mathcal{A}_{(2)} \in \mathbb{R}^{d_2 \times (d_1 d_3)}$ (if mode=2); or $\mathcal{A}_{(3)} \in \mathbb{R}^{d_3 \times (d_1 d_2)}$ (if mode=3)