Multi-Label Ranking Loss Minimization for Matrix Completion

Jiaxuan Li, Xiaoyan Zhu, Hongrui Wang, Yu Zhang, Xin Lai, Jiayin Wang

School of Computer Science and Technology, Xi'an Jiaotong University, China

AAAI-25

Background: Matrix Completion

Given an incomplete matrix $\boldsymbol{Y} \in \mathbb{R}^{n \times m}$, the completed matrix $\boldsymbol{X}$ should have two properties:

1) consistency: $\boldsymbol{X}$ should be as close as possible to $\boldsymbol{Y}$ in those entries that can be observed.

2) dependency: $\boldsymbol{X}$ should be in low-rank as $\boldsymbol{Y}$ can be completed only if it is redundant.

To summarize, the objective function can be formulated as:

$$\min_{\boldsymbol{X}} \lambda \|\boldsymbol{X}\|_* + \|\mathcal{R}_\Omega(\boldsymbol{X} - \boldsymbol{Y})\|_F^2$$

with

$$\mathcal{R}_\Omega(\boldsymbol{Y}_{ij}) = \begin{cases} \boldsymbol{Y}_{ij}, & (i,j) \in \Omega \\ 0, & (i,j) \notin \Omega \end{cases}$$

where $\Omega \subset \{(1, \cdots, n) \times (1, \cdots, m)\}$ is the index set of unobserved entries, $\|\boldsymbol{X}\|_*$ is the nuclear norm (a derivable surrogate of matrix rank), $\|\mathcal{R}_\Omega(\boldsymbol{X} - \boldsymbol{Y})\|_F^2$ is a surrogate for Hamming loss, λ is a balance parameter.

In real world applications, matrix may be measured by additional information, i.e., the features of customers and films for a film rating matrix. To consider such side information matrices $\boldsymbol{A} \in \mathbb{R}^{r_1 \times n}, \boldsymbol{C} \in \mathbb{R}^{r_2 \times m}$ for the rows and columns of $\boldsymbol{Y}$, the inductive mode of matrix completion can be summarized as:

$$\min_{\boldsymbol{X}} \lambda \|\boldsymbol{X}\|_* + \|\mathcal{R}_\Omega(\boldsymbol{A}\boldsymbol{X}\boldsymbol{C}^T - \boldsymbol{Y})\|_F^2$$

where $\boldsymbol{A}$ has r_1 features and $\boldsymbol{C}$ has r_2 features.

Furthermore, matrix factorization techniques can be taken to improve performance. For example, FNNM [1] takes $\boldsymbol{M} + \boldsymbol{A}\boldsymbol{N}\boldsymbol{C}^T$ to approximate the matrix to be completed:

$$\min_{\boldsymbol{X}} \lambda_1 \|\boldsymbol{M}\|_* + \lambda_2 \|\boldsymbol{N}\|_1 + \|\mathcal{R}_\Omega(\boldsymbol{M} + \boldsymbol{A}\boldsymbol{N}\boldsymbol{C}^T - \boldsymbol{Y})\|_F^2$$

where $\boldsymbol{M}$ aims at maintaining low-rank property and $\boldsymbol{N}$ is a sparse interaction matrix associating with the side information.

Motivation: Multi-Labe1 Learning

Multi-label learning deals with the problem where each instance is associated with a multiple labels.

That is, in multi-label learning, label matrix with multiple instances and multiple labels should be fitted and predicted.

Suppose that $\boldsymbol{X}$ is feature matrix and $\boldsymbol{Y}$ is label matrix, a common multi-label learning objective is:

$$\min_{\boldsymbol{W}} \|\boldsymbol{X}\boldsymbol{W} - \boldsymbol{Y}\|_F^2 + \lambda \|\boldsymbol{W}\|$$

where $\boldsymbol{W}$ is the coefficient matrix to be learned.

Apparently, multi-label techniques can be used to cope with matrix completion.

Multi-Label Ranking Loss (MLR)

Multi-label ranking loss evaluate the proportion of label pairs that are inversely ranked:

$$RL = \frac{1}{n} \sum_{i=1}^{n} \frac{\left|\{(p,q) | \boldsymbol{Y}_{ip} = 1, \boldsymbol{Y}_{iq} = -1, \widehat{\boldsymbol{Y}}_{ip} < \widehat{\boldsymbol{Y}}_{iq}\}\right|}{|\boldsymbol{Y}_i| \cdot |\overline{\boldsymbol{Y}}_i|}$$

where $\boldsymbol{Y} \in \{0,1\}^{n \times m}$ is the real label matrix and $\widehat{\boldsymbol{Y}}$ is the predicted one, $|\boldsymbol{Y}_i|, |\overline{\boldsymbol{Y}}_i|$ denote the number of positive and negative labels in i-th instance.

MLR for Matrix Completion

Advantages of MLR derive from its considering on the ranking relation between $\boldsymbol{Y}$ and $\widehat{\boldsymbol{Y}}$,

- It incorporates the relative relation in addition to linear relation among matrix vectors.
- It incorporates the pairwise correlation (local correlation, multi-label ranking loss minimization) in addition to matrix-wise correlation (global correlation, matrix rank minimization) and element-wise correlation (non correlation, Hamming loss minimization).

Weakness: It is hard to directly optimize the ranking loss between $\boldsymbol{Y}$ and $\widehat{\boldsymbol{Y}}$, due to different instances being differently labeled.

Relaxed MLR

To enable multi-label ranking loss minimization (MLRM), MLR is relaxed into the comparisons between every two columns:

$$\boldsymbol{R}_{ij} = \begin{cases} 1, & \boldsymbol{Y}_{ip} > \boldsymbol{Y}_{iq} \\ -1, & \boldsymbol{Y}_{ip} < \boldsymbol{Y}_{iq} \\ 0, & \boldsymbol{Y}_{ip} = \boldsymbol{Y}_{iq} \end{cases}$$

where $1 \leq p, q \leq c$ are column index in original matrix $\boldsymbol{Y}$, and $1 \leq j \leq \frac{c(c-1)}{2}$ is the column index in transformed matrix $\boldsymbol{R}$.

A more general method:

$$\boldsymbol{R}_{ij} = \boldsymbol{Y}_{ip} - \boldsymbol{Y}_{iq}$$

which can be formulated as:

$$\boldsymbol{R} = \boldsymbol{Y}\boldsymbol{L}$$

$$\boldsymbol{L} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ m-1 \\ m \end{array} \begin{array}{c} \begin{array}{ccccccccc} (1,2) & (1,3) & \cdots & (1,m) & (2,3) & \cdots & (2,m) & \cdots\cdots & (m-1,m) \end{array} \\ \begin{pmatrix} 1 & 1 & \cdots & 1 & & & & & \\ -1 & & & & 1 & \cdots & 1 & & \\ & -1 & & & -1 & & & & \\ & & \ddots & & & \ddots & & & \\ & & & & & & & & 1 \\ & & & -1 & & & -1 & & -1 \end{pmatrix} \end{array}$$

Method: MLRM

The matrix to be approximated can be shifted form $\boldsymbol{Y}$ to $\boldsymbol{R}$. MLRM essentially proposes a novel loss function for matrix completion can be expressed as several patterns, e.g.:

- transductive pattern

$$\min_{\boldsymbol{M}} \lambda \|\boldsymbol{M}\|_* + \frac{1}{2} \|\boldsymbol{M} - \boldsymbol{R}\|_F^2$$

- inductive pattern with matrix factorization:

$$\min_{\boldsymbol{M}} \lambda_1 \|\boldsymbol{M}\|_* + \lambda_2 \|\boldsymbol{N}\|_1 + \frac{1}{2} \|\boldsymbol{M} + \boldsymbol{A}\boldsymbol{N}\boldsymbol{B}^T - \boldsymbol{R}\|_F^2$$

where $\boldsymbol{B} = \boldsymbol{L}^T\boldsymbol{C}$

- inductive pattern with matrix factorization in global pattern:

$$\min_{\boldsymbol{M}} \lambda_1 \|\boldsymbol{M}\|_* + \lambda_2 \|\boldsymbol{N}\|_1 + \frac{1}{2} \|\boldsymbol{A}\boldsymbol{N}\boldsymbol{B}^T - \boldsymbol{M}\|_F^2 + \frac{1}{2} \|\boldsymbol{M} - \boldsymbol{R}\|_F^2$$

Above objectives can be solved by alternating minimization [2] as Alg. 1, and the final prediction will be:

$$\widehat{\boldsymbol{Y}} = (\widehat{\boldsymbol{M}} + \boldsymbol{A}\widehat{\boldsymbol{N}}\boldsymbol{B}^T)\boldsymbol{L}^-$$

where $\boldsymbol{L}^-$ can be viewed as the pseudoinverse of $\boldsymbol{L}$.

MLRM converts the incomplete matrix $\boldsymbol{Y}$ into the pairwise matrix $\boldsymbol{R} = \boldsymbol{Y}\boldsymbol{L}$, and thus is able to avoid the handle on function $\mathcal{R}_\Omega$. As a result, MLRM obtains longer strides and fewer iterations.

Algorithm 1: Framework of MLRM.

Input: Side matrices $\boldsymbol{A}, \boldsymbol{C}$, the matrix to be completed $\boldsymbol{Y}$.
Parameter: Hyper-parameters λ_1, λ_2.
Output: $\boldsymbol{M}, \boldsymbol{N}$.

Calculate convert matrix $\boldsymbol{L}$.
Calculate pairwise ranking matrix $\boldsymbol{R} = \mathcal{R}_\Omega(\boldsymbol{Y})\boldsymbol{L}$.
Update right side information matrix $\boldsymbol{B} = \boldsymbol{L}^T\boldsymbol{C}$.
while not converged **do**
  $\boldsymbol{E}_k \leftarrow \boldsymbol{R} - \boldsymbol{A}\boldsymbol{N}_k\boldsymbol{B}^T$;
  $\boldsymbol{M}_{k+1} \leftarrow \mathcal{D}_{\lambda_1}(\boldsymbol{E}_k)$;
  $\boldsymbol{F}_k \leftarrow \boldsymbol{N}_k - \frac{1}{L_p}\boldsymbol{A}^T(\boldsymbol{M}_{k+1} + \boldsymbol{A}\boldsymbol{N}_k\boldsymbol{B}^T - \boldsymbol{R})\boldsymbol{B}$;
  $\boldsymbol{N}_{k+1} \leftarrow \mathcal{S}_{\lambda_2}(\boldsymbol{F}_k)$;
  $k \leftarrow k + 1$;
end while
$\boldsymbol{M} \leftarrow \boldsymbol{M}_k, \boldsymbol{N} \leftarrow \boldsymbol{N}_k$;
return $\boldsymbol{M}, \boldsymbol{N}$.

References:

[1] Yang, M.; Li, Y.; and Wang, J. 2020. Feature and nuclear norm minimization for matrix completion. IEEE Transactions on Knowledge and Data Engineering, 34(5): 2190 2199.

[2] Beck, A. 2015. On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes. SIAM Journal on Optimization, 25(1): 185–209.

Codes: https://github.com/JiaxuanGood/MLRM.git

Email: lijiaxuan@stu.xjtu.edu.cn