Transductive Matrix Completion with Calibration for Multi-Task Learning

Hengfang Wang [1] Yasi Zhang [2] Xiaojun Mao [3] Zhonglei Wang [4]
[1] Fujian Normal University [2] UCLA [3] Shanghai Jiao Tong University [4] Xiamen University

Background

- Multi-task learning (MTL, Caruana, 1997) implements a robust learner for multiple tasks incorporating multiple sources, and it is commonly used in practice:
  - web search,
  - medical diagnosis,
  - natural language processing,
  - computer version.
- In MTL problems, features related to the responses may exist.
- However, it is inevitable that such features also suffer from missingness.
- External data sources have become available recently:
  - Summary statistics from census,
  - Summary information from other study,
  - ......
- Incorporating external information may improve estimation efficiency.
- However, seldom is done by calibration (Deville and Särndal, 1992) in the area of matrix completion.

Contribution

- Propose a Transductive Matrix Completion with Calibration (TMCC) algorithm to achieve the following goals:
  - Complete the feature and task matrices symultaneously.
  - Incorporate summary information by calibration.
- Theorecal properites:
  - Convergence rate is $O(k^{-2})$.
  - Constant order improvement is achieved using calibration.

Notations

- S: number of different types of task matrices.
- For $s \in [S]$,
  - $Y^{(s)} = (y_{ij}^{(s)}) \in \mathbb{R}^{n \times m_s}$: a task matrix suffers from missingness.
  - $Z_\star^{(s)} = (z_{\star,ij}^{(s)}) \in \mathbb{R}^{n \times m_s}$: low-rank distribution parameter such that

$$f^{(s)}(y_{ij}^{(s)} | z_{\star,ij}^{(s)}) = h^{(s)}(y_{ij}^{(s)}) \exp\{y_{ij}^{(s)} z_{\star,ij}^{(s)} - g^{(s)}(z_{\star,ij}^{(s)})\},$$

    - $h^{(s)}$ and $g^{(s)}$ are the base function and the link function, respectively.
  - $R_y^{(s)} = (r_{y,ij}^{(s)}) \in \mathbb{R}^{n \times m_s}$: indicator matrix for $Y^{(s)}$.
- $X = (x_{ij}) \in \mathbb{R}^{n \times d}$: noisy feature matrix suffers from missingness.
- $X_\star = (x_{\star,ij}) \in \mathbb{R}^{n \times d}$: low-rank true feature matrix such that $X = X_\star + \epsilon$.
- $R_x = (r_{x,ij}) \in \mathbb{R}^{n \times d}$: indicator matrix for X.
- A and B: available external summary information such that $AX_\star = B$.

Illustration

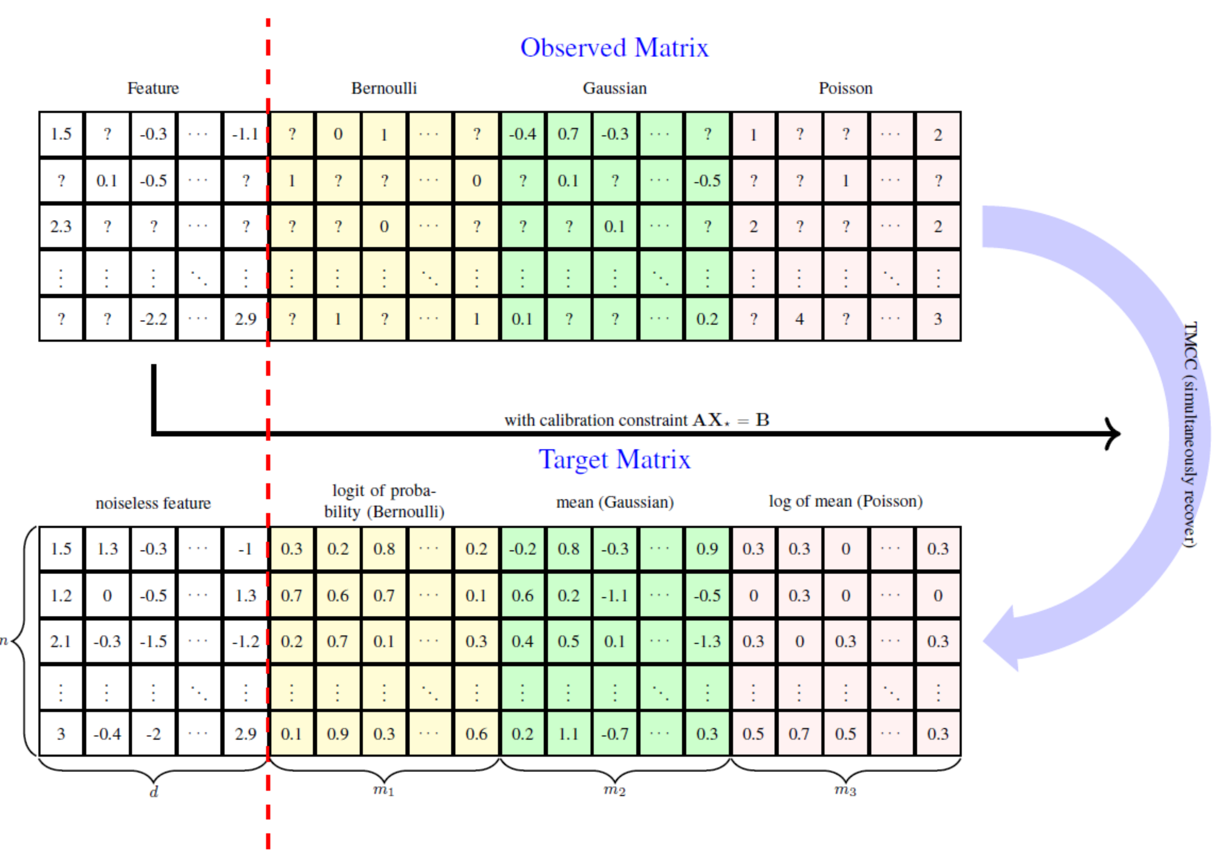

Figure 1: Algorithm Illustration.

Objective function

$$\begin{aligned} \widehat{M} &= \arg\min_{M^\dagger \in \mathbb{R}^{n \times D}} \Big[\frac{1}{nD} \Big\{ \ell(Z^\dagger) + \frac{1}{2} \| R_x \circ (X^\dagger - X) \|_F^2 \Big\} \\ &\qquad + \tau_1 \| AX^\dagger - B \|_F^2 + \tau_2 \| M^\dagger \|_\star \Big] \\ &=: \arg\min_{M^\dagger \in \mathbb{R}^{n \times D}} f_{\tau_1}(M^\dagger) + \tau_2 \| M^\dagger \|_\star \\ &=: \arg\min_{M^\dagger \in \mathbb{R}^{n \times D}} \mathcal{L}_{\tau_1, \tau_2}(M^\dagger), \end{aligned}$$

- $\ell(Z)$: negative quasi log-likelihood function,
- $M^\dagger = [X^\dagger, Z^\dagger]$,
- $\widehat{M} = [\widehat{X}, \widehat{Z}]$.

Algorithm

Algorithm 1: TMCC algorithm

Input: Incomplete matrices X, Y; indicator matrices R_x, R_y; calibration constraint matrices A and B, tuning parameters τ_1, τ_2; learning depth K, step size η, stopping criterion κ.
Initialize: Random matrices $M^{(0)} = M^{(1)} \in \mathbb{R}^{n \times D}$, $c = 1$.

```
for k = 1 to K do
    Compute θ = (c − 1)/(c + 2).
    Compute Q = (1 + θ)M^(k) − θM^(k−1).
    Compute T = Q − η∂f_τ1(Q).
    Compute M^(k+1) = T_ητ2(T).
    if L_τ1,τ2(M^(k+1)) > L_τ1,τ2(M^(k)) then
        c = 1;
    else
        c = c + 1;
    if |L_τ1,τ2(M^(k+1)) − L_τ1,τ2(M^(k))| ⩽ κ then
        M^‡ = M^(k+1);
        break;
```

Output: $M^\ddagger$.

Theoretical properties

Theorem

Under regularity conditions, we have

$$f_{\tau_1}\left(M^{(k)}\right) - f_{\tau_1}(M_\star) \leqslant \frac{2\tilde{L} \| M^{(0)} - M_\star \|_F^2}{\eta (k+1)^2},$$

with probability at least $1 - 4/(n+D)$, where $\tilde{L}$ is a constant related with other parameters.

Theorem

Under regularity conditions, with probability as least $1 - 4/(n+D)$,

$$\begin{aligned} &\left\{ \frac{1}{nD} + \frac{8\tau_1}{\tilde{L}_\alpha p_{\min}} \sigma_{\min}^2(A) \right\} \left\| \widehat{X} - X_\star \right\|_F^2 + \frac{1}{nD} \left\| \widehat{Z} - Z_\star \right\|_F^2 \\ &\leqslant \frac{crank(M_\star)}{nDp_{\min}^2} \left\{ \alpha^2 + \frac{\tilde{U}_\alpha \vee 1/\delta^2}{\tilde{L}_\alpha^2} \right\} \{ \gamma + \log^3(n \vee D) \}, \end{aligned}$$

where c, c_1, c_2 are positive constants.

- Under the assumption that the feature matrix is also regarded as a response matrix from Gaussian noise with unit variance, by comparing Theorem 7 in Apaya and Klopp (2019), we have a constant order improvement with the help of calibration infromation.
- Under generality condition, we have

$$\begin{aligned} &\frac{1}{nD} \left\| \widehat{X} - X_\star \right\|_F^2 + \frac{1}{nD} \left\| \widehat{Z} - Z_\star \right\|_F^2 \\ &\overset{(i)}{<} \left\{ \frac{1}{nD} + \frac{8\tau_1}{\tilde{L}_\alpha p_{\min}} \sigma_{\min}^2(A) \right\} \left\| \widehat{X} - X_\star \right\|_F^2 + \frac{1}{nD} \left\| \widehat{Z} - Z_\star \right\|_F^2, \end{aligned}$$

  where the inequality (i) is strict, which is one of the main theoretical contributions of this paper.

Experiments

- Setup:
  - $X_\star = X_\star^0 / \|X_\star^0\|_\infty$,
    - $X_\star^0 = PQ^T$,
    - $P \in \mathbb{R}^{n \times r}$ and $Q \in \mathbb{R}^{d \times r}$.
  - $Z_\star^{(s)} = \widetilde{Z}_\star^{(s)} / \|\widetilde{Z}_\star^{(s)}\|_\infty$ for $s = 1, 2, 3$,
    - Linear case:
    - $\widetilde{Z}_\star^{(s)} = X_\star W^{(s)}$,
    - $W^{(s)} \in \mathbb{R}^{d \times m_s}$.
    - Nonlinear case:
    - $z_{\star,ij}^{(s)} = t^{(s)}(x_{\star,ij})$,
    - $t^{(1)}(x) = x^2 + x + 0.5$,
    - $t^{(2)}(x) = -x^2 - x$,
    - $t^{(3)}(x) = -x^2 - 2x + 0.2$.
  - For the target matrices,
    - $Y^{(1)}$ has Bernoulli entries with support $\{0, 1\}$,
    - $Y^{(2)}$ has Poisson entries,
    - $Y^{(3)}$ has Gaussian entries with known $\sigma^2 = 1$.
  - For the external information, set $A = (1/n)\mathbf{1}_{1 \times n}$ and $B = AX_\star$.
  - $n = 1500$, $d = 500$, $m_1 = m_2 = m_3 = 500$.
- Algorithms:
  - MC_0: Same as TMCC, but without calibration.
  - CMC_SI: Collective matrix completion (Alaya and Klopp, 2019) is used to complete the parameters for the target matrix, and Soft-Impute method from (Mazumder et al., 2010) is used to complete the feature matrix separately.
  - TS: A two-stage method, where, at the first stage, only the feature matrix is imputed by the Soft-Impute method, and at the second stage, the method MC_0 is applied to the concatenated matrix joined by the feature matrix and the observed response matrices.
- Evaluation creiteria:
  - Relative error (RE):

$$\mathrm{RE}(\widehat{M}) = \|\widehat{M} - M_\star\|_F / \|M_\star\|_F,$$

    - $M_\star$: a certain target matrix suffering from missingness,
    - $\widehat{M}$: completed matrix.

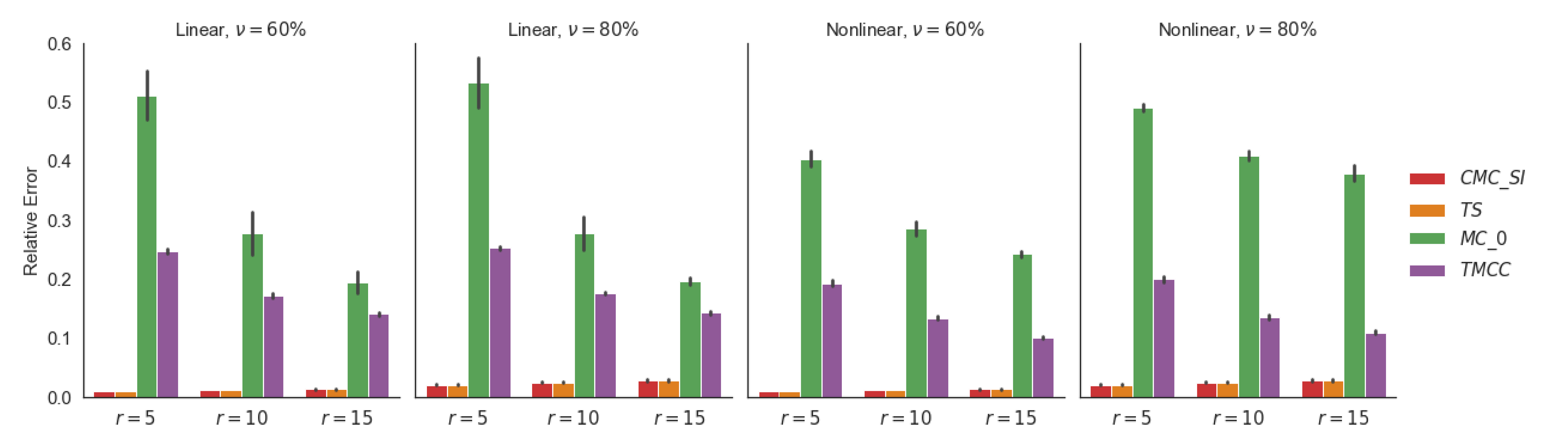

Figure 2: Relative Error of Feature Matrix (with the Black Lines Representing Â±theStandardError)

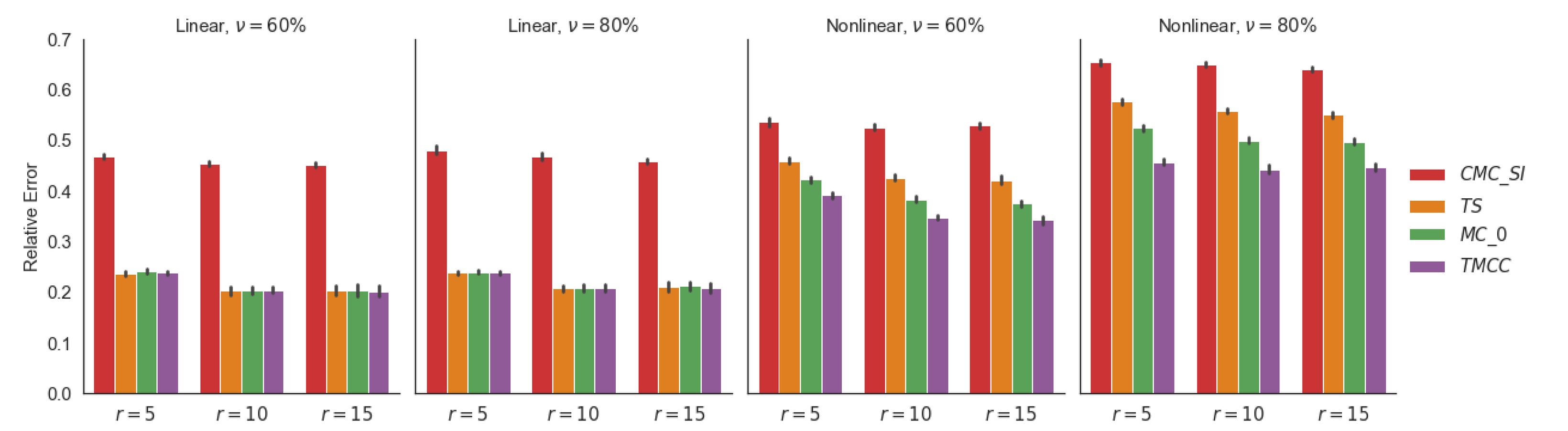

Figure 3: Relative Error of Target Matrix (with the Black Lines Representing Â±theStandardError)

Reference

- Alaya, M. Z. and O. Klopp (2019). Collective matrix completion. Journal of Machine Learning Research 20 (148), 1-43.
- Caruana, Rich (1997), Multitask learning, Machine Learning, 28(1), 41â€“75.
- Deville, J.-C. and C.-E. Särndal (1992). Calibration estimators in survey sampling. Journal of the American Statistical Association 87 (418), 376-382.
- Mazumder, R., T. Hastie, and R. Tibshirani (2010). Spectral regularization algorithms for learning large incomplete matrices. Journal of Machine Learning Research 11 (Aug), 2287-2322.