Convolutional Imputation of Matrix Networks

Qingyun Sun Mengyuan Yan
David Donoho Stephen Boyd

Presenter: Yevgeny Tkach

2019 Spring @
https://qdata.github.io/deep2Read/

Executive Summary

- Graph G=(V,E) where every node v, is represented by a matrix M.
- Matrices are only partially observed and the task is to perform matrices completion by leveraging extra information provided by G.
- Authors approach is using SVD for matrix completion and using spectral theory to propagate information from the graph.
- Experiments are essentially synthetical or consider sequential frames.

Outline

- Matrix completion problem and SVD
- Matrix network
- Convolutional Imputation Algorithm
- Experiments
- Conclusions

Matrix completion problem

- Consider a matrix X with partially observed entries from Ω.
- We notate $\mathcal{P}_\Omega : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ which defined $P_\Omega(X) = \begin{cases} X_{ij}, & (i,j)\in\Omega \\ 0, & otherwise \end{cases}$
- The problem is finding the best matrix M that solves

$$\begin{aligned} &\min_M \quad rank(M) \\ &s.t. \quad P_\Omega(M) = P_\Omega(X) \end{aligned}$$

- "The program is a common sense approach which simply seeks the simplest explanation fitting the observed data"
- This NP-hard Non-Convex Problem is usually relaxed to:

$$\begin{aligned} &\min_M \quad \|M\|_* \\ &s.t. \quad P_\Omega(M) = P_\Omega(X) \end{aligned}$$

$\|M\|_* = \sum_{k=1}^{n} \sigma_k(M)$, where $\sigma_k(M)$ is the k-th largest singular value of M

- This is usually solved with Lagrange multipliers which also allow noise

$$\underset{M}{\text{minimize}} \quad H(M) := \quad \tfrac{1}{2}\|P_\Omega(X-M)\|_F^2 + \lambda\|M\|_*$$

Low Rank SVD

- For low rank matrices solving the problem becomes feasible and most research is focused on such matrices.
- The following Theorem is the theoretic foundation for popular low rank SVD algorithms:

Theorem 1 *Let $X_{m\times n}$ be a matrix (fully observed), and let $0 < r \le \min(m,n)$. Consider the optimization problem*

$$\underset{Z:\, \operatorname{rank}(Z)\le r}{\text{minimize}} \; F_\lambda(Z) := \frac{1}{2}||X - Z||_F^2 + \lambda||Z||_*. \tag{9}$$

A solution is given by

$$\widehat{Z} \;=\; U_r \mathcal{S}_\lambda(D_r) V_r^T, \tag{10}$$

where the rank-r SVD of X is $U_r D_r V_r^T$ and $\mathcal{S}_\lambda(D_r) = diag[(\sigma_1 - \lambda)_+, \ldots, (\sigma_r - \lambda)_+]$.

Low Rank SVD

- The SVD solution is often calculated sequentially until some convergence criteria:

1. Replace the missing entries in X with the corresponding entries from the current estimate $\widehat{M}$:
$$\hat{X} \leftarrow P_\Omega(X) + P_\Omega^\perp(\widehat{M}); \tag{2}$$

2. Update $\widehat{M}$ by computing the soft-thresholded SVD of $\hat{X}$:
$$\hat{X} = UDV^T \tag{3}$$
$$\widehat{M} \leftarrow U\mathcal{S}_\lambda(D)V^T, \tag{4}$$

- An important part of the low rank SVD approximation theory is specifically defining the assumptions on the available data such that a “good” completion can be guaranteed.
- Specifically it requires the number of samples to be proportional to the Incoherence (“spread out”) of the matrix.

Matrix Network

- Our problem is to complete N different matrices where different matrices are connected between each other.
- The connection between the matrices represented via graph G=(V,E) with adjacency matrix W.
- matrix network A maps each node k in the graph to it's matrix A(k).
- The matrix completion problem is defined in the Furrier domain (Spectral Theory):
- U is a unitary N×N matrix, and the eigenvectors of L, the normalized Laplacian matrix of W, are the row vectors of U.
- The graph Furrier transform of A, $\hat{A} = \mathcal{U}A$, $\hat{A}(k) = \sum_{i \in J} U(k,i)A(i)$
- The convex optimization target:

$$L_\lambda(\hat{M}) = \frac{1}{2}\|A^\Omega - P_\Omega \mathcal{U}^* \hat{M}\|_2^2 + \sum_{k=1}^{N} \lambda_k \|\hat{M}(k)\|_*$$

Convolutional Imputation Algorithm

Iterative Imputation:

input $P_\Omega(A)$.

Initialization $A_0^{\text{est}} = 0, t = 0$.

for $\lambda^1 > \lambda^2 > \ldots > \lambda^C$, where $\lambda^j = (\lambda_k^j), k = 1, \ldots, N$ **do**

repeat

$A^{\text{impute}} = P_\Omega(A) + P_\Omega^\perp(A_t^{\text{est}})$.

$\hat{A}^{\text{impute}} = \mathcal{U} A^{\text{impute}}$.

$\hat{A}_{t+1}^{\text{est}}(k) = S_{\lambda_k^j}(\hat{A}^{\text{impute}}(k))$.

$A_{t+1}^{\text{est}} = \mathcal{U}^{-1} \hat{A}_{t+1}^{\text{est}}$.

t=t+1.

until $\|A_t^{\text{est}} - A_{t-1}^{\text{est}}\|^2 / \|A_{t-1}^{\text{est}}\|^2 < \epsilon$.

Assign $A_{\lambda^j} = A_t^{\text{est}}$.

end for

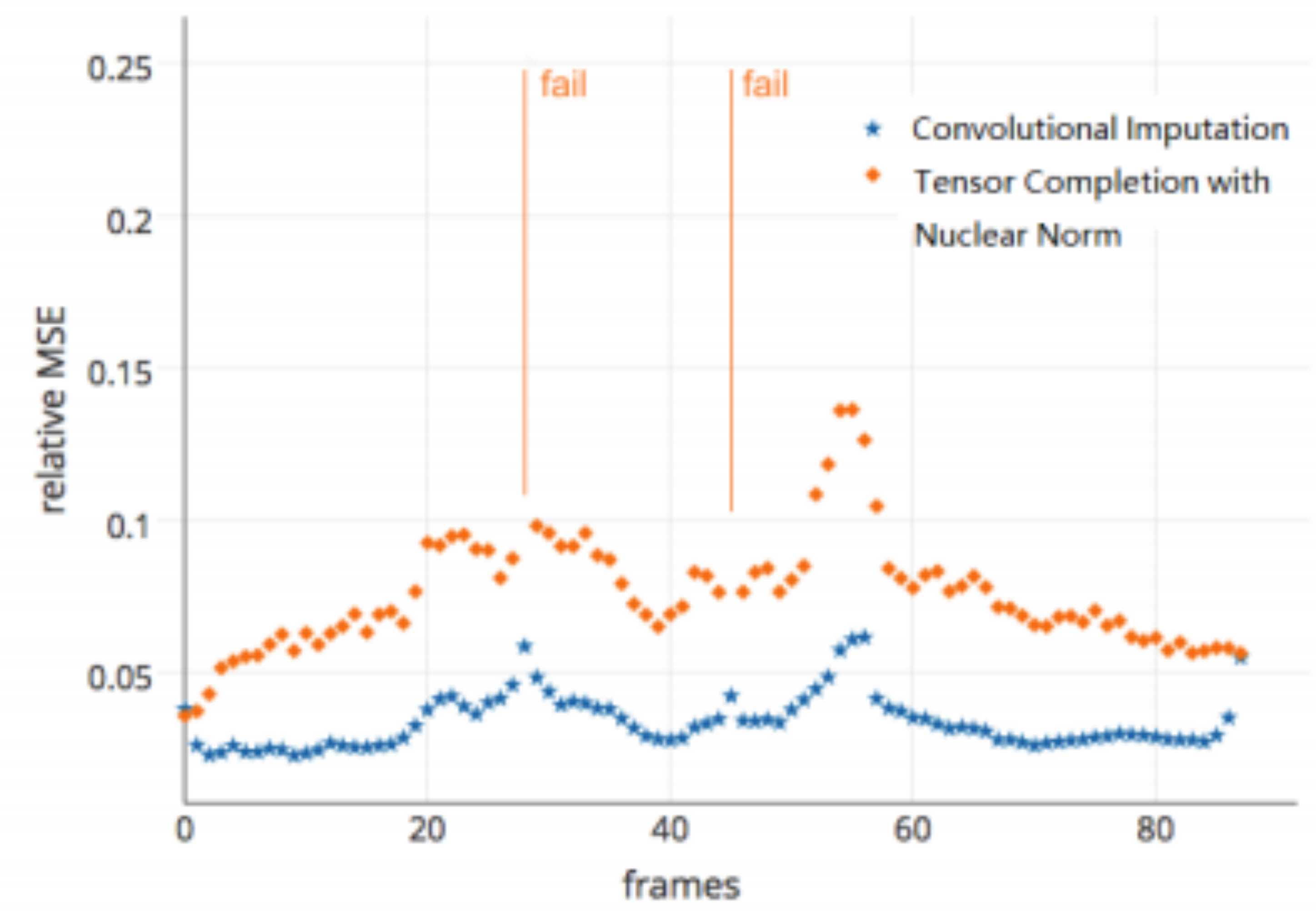
output The sequence of solutions $A_{\lambda^1}, \ldots, A_{\lambda^C}$.

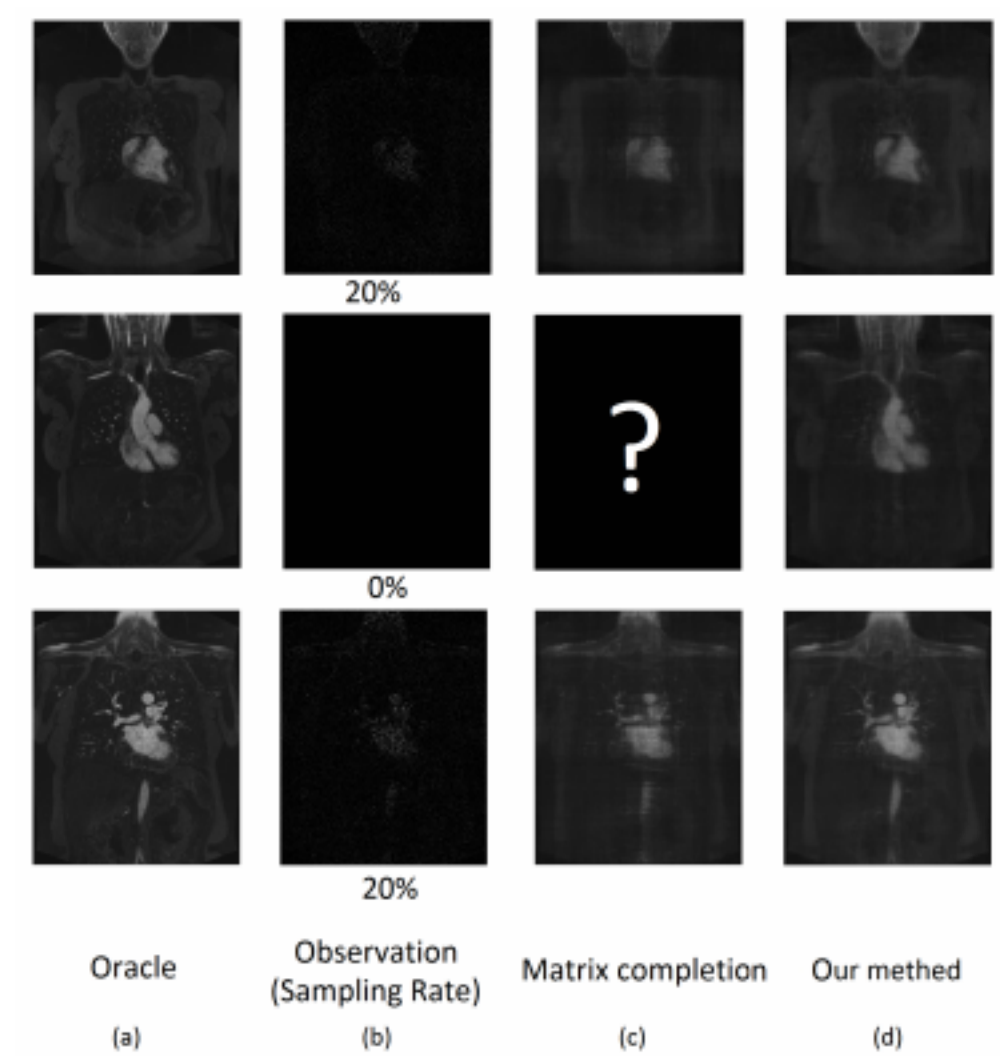
- Where $S_\lambda(\hat{A}) = V_1(\Sigma - \lambda \mathrm{I})_+ V_2^*$ and $\hat{A} = V_1 \Sigma V_2^*$
- For this algorithm convergence is guaranteed under coherence and sampling assumptions

Experiments

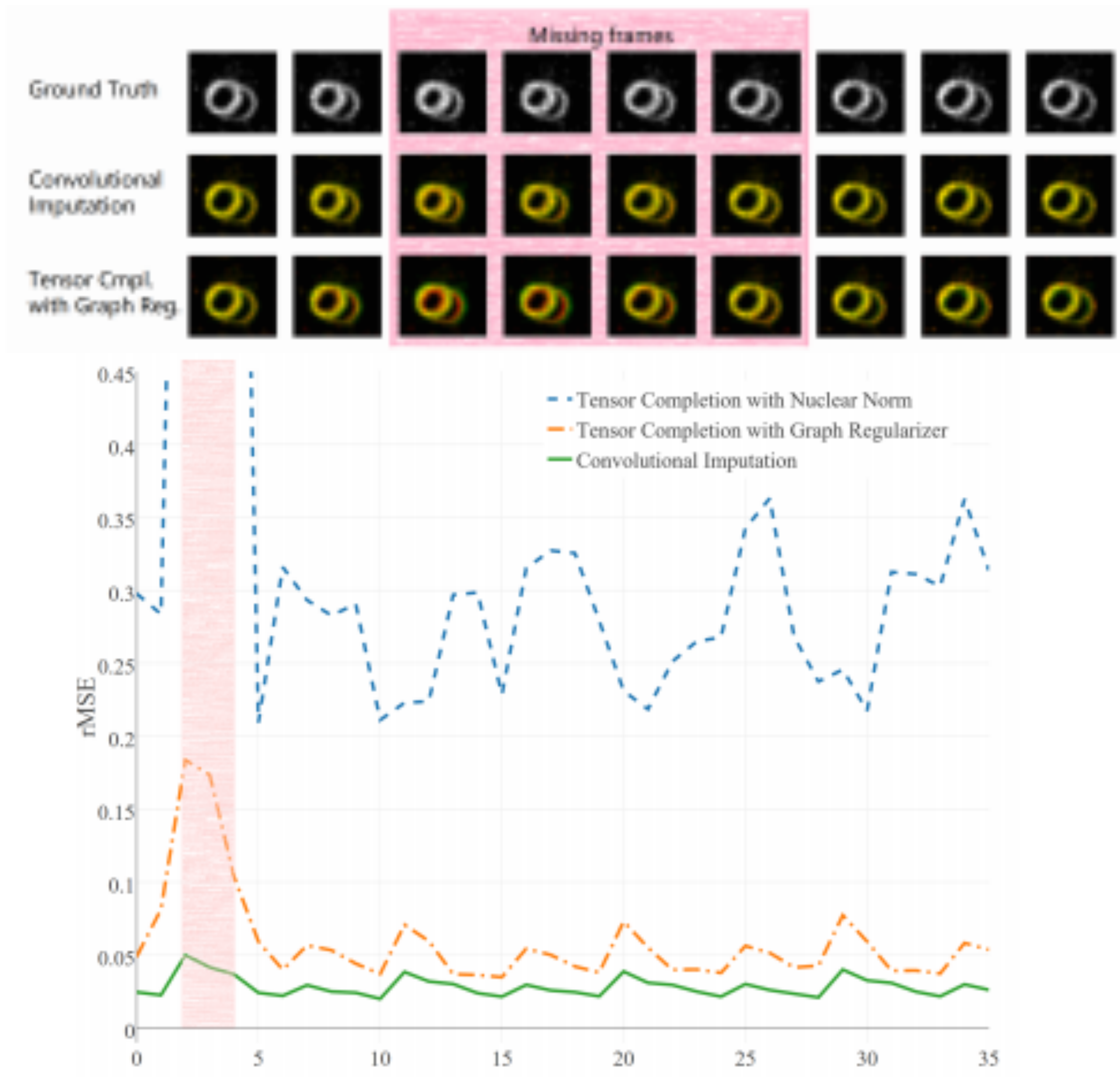
- **Feature matrices on Facebook network -** This is a synthetic dataset that is based on a real sample from the Facebook network. The feature matrices (the low rank SVD) for each node were randomly generated, and the purpose of the experiment was to check that these matrices cn be recovered
- **MRI completion -** In the 88 frames there are 2 frames missing, and pixels sampled from the rest i.i.d. from a Bernoulli distribution with p = 0.2.
- **SPECT completion -** The sequence has 36 frames, capturing 4 periods of heart beats. 4 consecutive frames out of the 36 frames are missing and the other frames are sampled i.i.d. from a Bernoulli distribution with p = 0.2

MRI completion





SPECT completion



Discussion

- The problem is novel, however we don’t seem to have real world data sets that capture the true nature of the problem
- Other approaches that this model compares to were not designed for the problem so the comparison is not that interesting
- Their result is quite a natural extension to the existing SVD based matrix completion literature