icassp 2022 Singapore

APPLYING DIFFERENTIAL PRIVACY TO TENSOR COMPLETION

Zheng Wei[1], Zhengpin Li[1], Xiaojun Mao[2] and Jian Wang[1†]

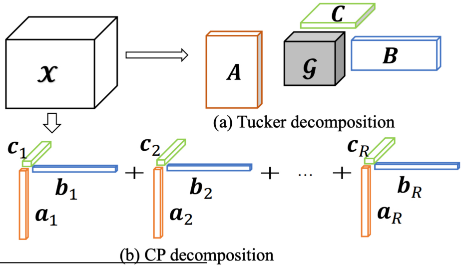
[1]School of Data Science, Fudan University, China
[2]School of Mathematical Sciences, Shanghai Jiao Tong University, China

†Z. Wei and Z. Li contributed equally to this work. X. Mao and J. Wang are corresponding authors (E-mail: maoxj@sjtu.edu.cn; jian_wang@fudan.edu.cn).

Outline

Tensor Completion

- Tensor completion aims at filling the missing or unobserved entries based on partially observed tensors.
- Two most widely used tensor decomposition methods:
  1. CANDECOMP/PARAFAC (CP) decomposition[1]
  2. Tucker decomposition[2]



[1] Frank L Hitchcock. “The expression of a tensor or a polyadic as a sum of products”. In: *Journal of Mathematics and Physics* 6.1-4 (1927), pp. 164–189.

[2] Ledyard R Tucker. “Some mathematical notes on three-mode factor analysis”. In: *Psychometrika* 31.3 (1966), pp. 279–311.

Differential Privacy

Definition

A (randomized) algorithm $\mathcal{A}$ whose outputs lie in a domain $\mathcal{S}$ is said to be ϵ-differentially private if for all subsets $S \subseteq \mathcal{S}$, for all datasets $\mathcal{D}$ and $\mathcal{D}'$ that differ in at most one entry, it holds that:

$$\Pr(\mathcal{A}(\mathcal{D}) \in S) \leq e^{\epsilon} \Pr\left(\mathcal{A}\left(\mathcal{D}'\right) \in S\right). \quad (1)$$

- Generally by adding noise to solve this problem[3]
- ϵ represents the level of privacy protection (privacy budget)

[3] Cynthia Dwork et al. "Calibrating noise to sensitivity in private data analysis". In: *Theory of cryptography conference*. Springer. 2006, pp. 265–284.

Backbone Algorithm

Tucker decomposition for a tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ is:

$$\mathcal{X} \approx \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{t=1}^{T} g_{pqt} \mathbf{a}_{p:} \circ \mathbf{b}_{q:} \circ \mathbf{c}_{t:} \tag{2}$$

If $\mathcal{G}$ is superdiagonal (i.e., $g_{pqr} \equiv g_r$) and $P = Q = R$, Tucker decomposition would reduce to CP decomposition:

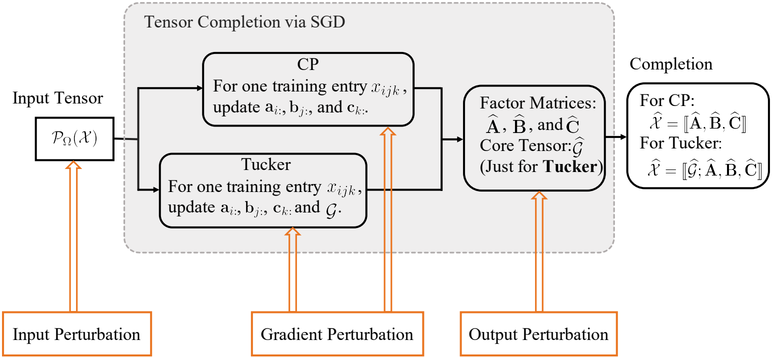
$$\mathcal{X} \approx \sum_{r=1}^{R} g_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r = [\![\mathbf{g}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \tag{3}$$

By imposing a F-norm penalty to restrict the complexity of the core tensor, the Tucker decomposition problem can be reformulated as:

$$\begin{aligned} \min_{\mathbf{A},\mathbf{C},\mathcal{G}} \quad & f(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathcal{G}) = \|\mathcal{P}_\Omega(\mathcal{X} - [\![\mathcal{G}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!])\|_F^2 \\ & + \lambda_o \left(\|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2 + \|\mathbf{C}\|_F^2\right) + \lambda_g \|\mathcal{G}\|_F^2 \end{aligned} \tag{4}$$

Overall

Various perturbation approaches within tensor completion framework.



Input Perturbation

Algorithm 1 Private Input Perturbation

Input: $\mathcal{X}$: noisy incomplete tensor, Ω: indexes set of observations, d: rank of tensor, λ_o: regularization parameter for the factor matrices, λ_g: regularization parameter for the core tensor, ϵ: privacy budget

1: Generate each entry of noise tensor $\mathcal{N}$ by $\mathrm{Lap}(\Delta_{\mathcal{X}}^{(l)}/\epsilon)$

2: Let $\mathcal{X}' = \{x_{ijk} + n_{ijk} | (i,j,k) \in \Omega\}$

3: Use $\mathcal{X}'$ as input to solve Tucker decomposition via SGD and obtain estimated $\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \widehat{\mathbf{C}}$ and $\widehat{\mathcal{G}}$

Output: Estimated $\widehat{\mathbf{A}} \in \mathbb{R}^{n_1 \times d}$, $\widehat{\mathbf{B}} \in \mathbb{R}^{n_2 \times d}$, $\widehat{\mathbf{C}} \in \mathbb{R}^{n_3 \times d}$ and $\widehat{\mathcal{G}} \in \mathbb{R}^{d \times d \times d}$

Gradient Perturbation

Algorithm 2 Private Gradient Perturbation

Input: n: number of iterations, ϵ: privacy budget, η: learning rate, m: clipping constant

Initialize random factor matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathcal{G}$

for n iterations **do**

  for $x_{ijk} \in \mathcal{X}$ **do**

    $\mathbf{a}_{i:} \leftarrow \mathbf{a}_{i:} - \eta \nabla_{\mathbf{a}_{i:}} f$

    $\mathbf{b}_{j:} \leftarrow \mathbf{b}_{j:} - \eta \nabla_{\mathbf{b}_{j:}} f$

    $\nabla_{\mathbf{c}_{k:}} f \leftarrow \nabla_{\mathbf{c}_{k:}} f \,/\, \max\left(1, \|\nabla_{\mathbf{c}_{k:}} f\|_2 / m\right)$

    Sample noise vector $\mathbf{n}_{i:}$ satisfying $p(\mathbf{n}_{i:}) \propto exp\{-\frac{\varepsilon\|\mathbf{n}_{i:}\|}{\Delta_{\mathcal{X}}^{(G)}}\}$

    $\mathbf{c}_{k:} \leftarrow \mathbf{c}_{k:} - \eta(\nabla_{\mathbf{c}_{k:}} f + \mathbf{n}_{i:})$

    $\mathcal{G} \leftarrow \mathcal{G} - \eta \nabla_{\mathcal{G}} f$

  end for

end for

Output Perturbation

Algorithm 3 Private Output Perturbation

Input: $\mathcal{X}$: noisy incomplete tensor, Ω: indexes set of observations, d: rank of tensor, ϵ: privacy budget

Solve Tucker decomposition via SGD and obtain estimated $\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \widehat{\mathbf{C}}$ and $\widehat{\mathcal{G}}$

Sample noise matrix $\mathbf{N}$, all rows of which are sampled from $\exp\left\{-\frac{\epsilon\|\mathbf{n}_{i:}\|}{\Delta_{\mathcal{X}}^{(O)}}\right\}$

3: $\widehat{\mathbf{C}} \leftarrow \widehat{\mathbf{C}} + \mathbf{N}$

Output: Estimated $\widehat{\mathbf{A}} \in \mathbb{R}^{n_1 \times d}$, $\widehat{\mathbf{B}} \in \mathbb{R}^{n_2 \times d}$, $\widehat{\mathbf{C}} \in \mathbb{R}^{n_3 \times d}$ and $\widehat{\mathcal{G}} \in \mathbb{R}^{d \times d \times d}$

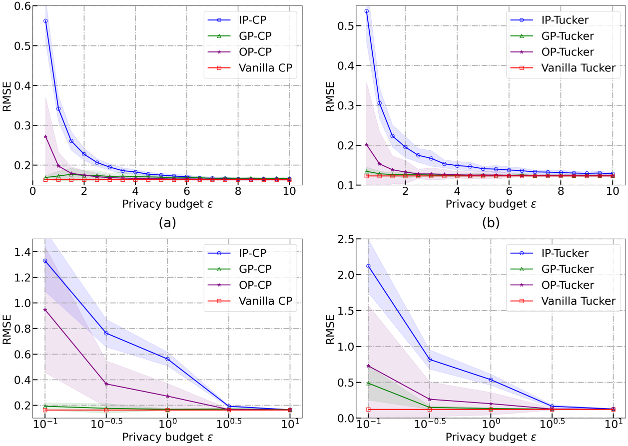
Synthetic Dataset

We set the size and rank of $\mathcal{X}$ to $20 \times 20 \times 20$ and 3, respectively.

- For CP decomposition: $\mathcal{X} = [\![\widetilde{\mathbf{A}}, \widetilde{\mathbf{B}}, \widetilde{\mathbf{C}}]\!] + \mathcal{N}$ where $\widetilde{\mathbf{A}} \in \mathbb{R}^{20\times 3}$, $\widetilde{\mathbf{B}} \in \mathbb{R}^{20\times 3}$ and $\widetilde{\mathbf{C}} \in \mathbb{R}^{20\times 3}$ are from standard normal distribution, and $\mathcal{N}$ represents a mean zero Gaussian noise tensor satisfying that signal-to-noise (SNR) is one.
- For Tucker decomposition: We draw the entries of the core tensor $\widetilde{\mathcal{G}} \in \mathbb{R}^{3\times 3\times 3}$ from standard normal distribution and construct $\mathcal{X}$ via $\mathcal{X} = [\![\widetilde{\mathcal{G}}; \widetilde{\mathbf{A}}, \widetilde{\mathbf{B}}, \widetilde{\mathbf{C}}]\!] + \mathcal{N}$ where $\mathcal{N}$ is same as the generation in CP decomposition.

Synthetic Results

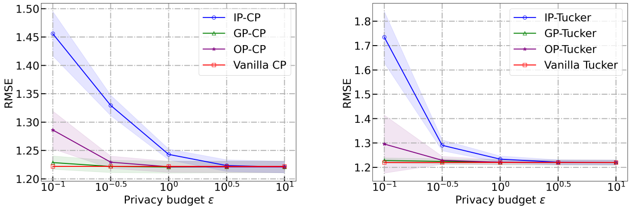
Performance comparison of CP and Tucker decompositions. The left and right columns present the performance of CP decomposition and Tucker decomposition, respectively.



(a) (b) (c) (d)

Empirical Results

We also study the performance of our methods on Movie-Lens 100K[4] datasets, which consists of 943 users, 1682 movies and 212 timestamps with density 6.30%.



(a) CP Decomposition (b) Tucker Decomposition

[4] F Maxwell Harper and Joseph A Konstan. "The movielens datasets: History and context". In: *Acm Transactions on interactive intelligent Systems (TiiS)* 5.4 (2015), pp. 1–19.

Conclusion

The contributions of our work are summarized as follows.

- We are the first to propose a solid and unified framework for applying differential privacy to tensor completion.
- We provide complete algorithm procedures and theoretical analysis for each privacy-preserving approach in our framework.
- Experimental results on synthetic and real-world datasets demonstrate that the proposed approaches can yield high accuracy, while ensuring strong privacy protections.

Thank you!