Appendix for "Federated Learning Empowered Neural Tensor Completion for Accurate IoT Data Recovery"

Chunsheng Liu

Abstract

This document is the appendix file for the IEEE INFOCOM 2023 paper titled "Federated Learning Empowered Neural Tensor Completion for Accurate IoT Data Recovery".

I. Proof of Matroid Base Constraint in Model (20)

Lemma A1: Given $\mathcal{M} := \{(s,u)|\forall s \in \mathcal{S}, u \in \mathcal{U}\}$, the pair $\mathcal{A} := \{\mathcal{M}, \mathcal{I}\}$ is a matroid, where $\mathcal{I}$ is a collection of independent sets, that is, $\mathcal{I} := \{\mathcal{V}|\mathcal{V} \subseteq \mathcal{G}, \forall v_1 = (s_1,u_1), v_2 = (s_2,u_2) \in \mathcal{V}, v_1 \neq v_2\}$. The constraint in $\mathbf{P1}$ corresponds to a matroid base constraint, *i.e.,* $\mathcal{V} \subseteq \mathcal{M}, \mathcal{V} \in B(\mathcal{A})$, where $B(\mathcal{A})$ is the set of bases of $\mathcal{A}$.

Proof A1: First, the nonempty property of $\mathcal{I}$ is obvious due to $U \geq 2$ and $S \geq 2$. Second, if $\mathcal{V}_1 \subseteq \mathcal{V}_2 \in \mathcal{I}$, then $\mathcal{V}_1 \in \mathcal{I}$. If not, there exist at least two different elements $v_1, v_2 \in \mathcal{V}_1$ that share the same second component. Since $\mathcal{V}_1 \subseteq \mathcal{V}_2$ holds, $v_1, v_2 \in \mathcal{V}_2$, which contradicts $\mathcal{V}_2 \in \mathcal{I}$. Lastly, suppose $\mathcal{V}_1, \mathcal{V}_2 \in \mathcal{I}$ and $|\mathcal{V}_1| < |\mathcal{V}_2|$, if there does not exist an element $v \in \mathcal{V}_2 \backslash \mathcal{V}_1$ such that $\mathcal{V}_1 \cup \{v\} \in \mathcal{I}$, then for any element $e \in \mathcal{V}_2 \backslash \mathcal{V}_1$, $\mathcal{V}_1 \cup \{v\} \notin \mathcal{I}$. Since $\mathcal{V}_1 \in \mathcal{I}$, each element in $\mathcal{V}_2 \backslash \mathcal{V}_1$ has the same UE in $\mathcal{V}_1$. Due to $|\mathcal{V}_1| < |\mathcal{V}_2|$, there exist at least two elements in $\mathcal{V}_2 \backslash \mathcal{V}_1$ with the same UE in $\mathcal{V}_1$. This implies the two elements sharing the same UE are also in $\mathcal{V}_2$ that contradicts with $\mathcal{V}_2 \in \mathcal{I}$. Therefore, $\mathcal{A} := \{\mathcal{M}, \mathcal{I}\}$ is a matroid.

According to the construction of $\mathcal{A}$, it is easy to obtain that the size of the matroid $\mathcal{A}$ is U because there are U UEs. Considering the constraint in $\mathbf{P1}$, $\forall u \in \mathcal{U}, \sum_{s \in \mathcal{S}} z_{su} = 1$ means finding an edge association strategy for every UE. It is equal to find a set $\mathcal{V} \subseteq \mathcal{M}$, which constitutes a base of $\mathcal{A}$, that is, $\mathcal{V} \in B(\mathcal{A})$. The lemma is thus proved. ■

II. Proof of the properties of $\tilde{J}$ in $\mathbf{P3}$

$$\begin{aligned} \tilde{J}_t(\mathcal{V}) &= \mu_t \tilde{T}|_{\forall (s,u)\in\mathcal{V}, z_{su}=1;\,\mathrm{otherwise}\, z_{su}=0} \\ &= \mu_t \max_{(s,u)\in\mathcal{V}} \{\gamma(\epsilon) t^s_{train})\} + 2\mu_t \max_{(s,u)\in\mathcal{V}} \{\tau_{su}\}|\mathcal{V}|, \end{aligned} \tag{1}$$

$$\begin{aligned} J_\ell(\mathcal{V}) &= \mu_\ell \mathcal{L}|_{\forall (s,u)\in\mathcal{V}, z_{su}=1;\, otherwise\, z_{su}=0} \\ &= \frac{\mu_\ell}{|\mathcal{S}|} \sum_{(s,u)\in\mathcal{V}} (||\mathcal{T}^s - \mathcal{X}^s||_F^2 + \sum_{k=1}^{n} \lambda_k^s ||\mathbf{A}_k^s||_F^2). \end{aligned} \tag{2}$$

Proof A2: According to the definition in Eq. (1) and (2), all coefficients in NTC-FL model are non-negative and then $\tilde{J}$ is nonnegative.

Since the expansion of any set $\mathcal{V} \subseteq \mathcal{M}$ will relax item $\mathcal{U}^s$ and increase the optimal objective value potentially. For example, when adding one element $v = (s_v, u_v)$ in $\mathcal{V}$, it is equal to associate UE u_v with edge server s_v, which possibly induces more loss and increases the system latency as well. Then, $\forall \mathcal{V}_1 \subseteq \mathcal{V}_2 \subseteq \mathcal{G}$, $\tilde{J}(\mathcal{V}_1) \leq \tilde{J}(\mathcal{V}_2)$ holds, that is, $\tilde{J}$ is monotone.

Next, to prove $\tilde{J}$ is supermodular, we just need prove $\tilde{J}_t(\mathcal{V})$ and $J_\ell(\mathcal{V})$ are supermodular.

According to the definition of super-modular, i.e., for a given finite ground set $\mathcal{M}$ and a real-valued set function defined as $J : 2^{\mathcal{M}} \to \mathbb{R}$, J is super-modular if and only if $J(\mathcal{V}_1) + J(\mathcal{V}_2) \leq J(\mathcal{V}_1 \cup \mathcal{V}_2) + J(\mathcal{V}_1 \cap \mathcal{V}_2)$ for $\forall \mathcal{V}_1, \mathcal{V}_2 \subseteq \mathcal{G}$. Then, for $J_\ell(\mathcal{V})$, we have

$$\begin{aligned} J_\ell(\mathcal{V}) &= \mu_\ell \mathcal{L}|_{\forall (s,u)\in\mathcal{V}, z_{su}=1;\, otherwise\, z_{su}=0} \\ &= \frac{\mu_\ell}{|\mathcal{S}|} \sum_{(s,u)\in\mathcal{V}} (||\mathcal{T}^s - \mathcal{X}^s||_F^2 + \sum_{k=1}^{n} \lambda_k^s ||\mathbf{A}_k^s||_F^2). \end{aligned} \tag{3}$$

For two given sets $\forall \mathcal{V}_1, \mathcal{V}_2 \subseteq \mathcal{G}$, let $J_\ell^l = J_\ell(\mathcal{V}_1) + J_\ell(\mathcal{V}_2)$ and $J_\ell^r = J_\ell(\mathcal{V}_1 \cup \mathcal{V}_2) + J_\ell(\mathcal{V}_1 \cap \mathcal{V}_2)$, then we have

$$\begin{aligned} J_\ell^l &= J_\ell(\mathcal{V}_1) + J_\ell(\mathcal{V}_2) \\ &= \mu_\ell/|\mathcal{S}|(\sum_{(s,u)\in\mathcal{V}_1} (||\mathcal{T}^s - \mathcal{X}^s||_F^2 + \sum_{k=1}^{n} \alpha_k^s ||\mathbf{A}_k^s||_F^2) + \sum_{(s,u)\in\mathcal{V}_2} (||\mathcal{T}^s - \mathcal{X}^s||_F^2 + \sum_{k=1}^{n} \alpha_k^s ||\mathbf{A}_k^s||_F^2)) \end{aligned} \tag{4}$$

$$\begin{aligned} J_\ell^r &= J_\ell(\mathcal{V}_1 \cup \mathcal{V}_2) + J(\mathcal{V}_\ell \cap \mathcal{V}_2) \\ &= \mu_\ell/|\mathcal{S}|(\sum_{(s,u)\in\mathcal{V}_1\cup\mathcal{V}_2} (||\mathcal{T}^s - \mathcal{X}^s||_F^2 + \sum_{k=1}^{n} \alpha_k^s ||\mathbf{A}_k^s||_F^2) + \sum_{(s,u)\in\mathcal{V}_1\cap\mathcal{V}_2} (||\mathcal{T}^s - \mathcal{X}^s||_F^2 + \sum_{k=1}^{n} \alpha_k^s ||\mathbf{A}_k^s||_F^2)) \end{aligned} \tag{5}$$

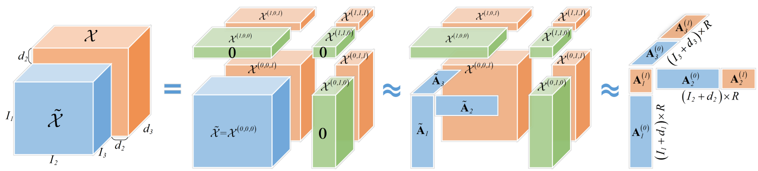

Fig. 1. An illustration of the partitioning property of CP decomposition.

Then,

$$\begin{aligned}
J_\ell^l - J_\ell^r &= J_\ell(\mathcal{V}_1) + J_\ell(\mathcal{V}_2) - (J_\ell(\mathcal{V}_1 \cup \mathcal{V}_2) + J_\ell(\mathcal{V}_1 \cap \mathcal{V}_2)) \\
&= \mu_\ell/|\mathcal{S}|(\sum_{(s,u)\in\mathcal{V}_1} (||\mathcal{T}^s - \mathcal{X}^s||_F^2 + \sum_{k=1}^{n} \alpha_k^s ||\mathbf{A}_k^s||_F^2) + \sum_{(s,u)\in\mathcal{V}_2} (||\mathcal{T}^s - \mathcal{X}^s||_F^2 + \sum_{k=1}^{n} \alpha_k^s ||\mathbf{A}_k^s||_F^2) \\
&\quad - \sum_{(s,u)\in\mathcal{V}_1\cup\mathcal{V}_2} (||\mathcal{T}^s - \mathcal{X}^s||_F^2 + \sum_{k=1}^{n} \alpha_k^s ||\mathbf{A}_k^s||_F^2) - \sum_{(s,u)\in\mathcal{V}_1\cap\mathcal{V}_2} (||\mathcal{T}^s - \mathcal{X}^s||_F^2 + \sum_{k=1}^{n} \alpha_k^s ||\mathbf{A}_k^s||_F^2)) \\
&= \mu_\ell/|\mathcal{S}|(\sum_{(s,u)\in\mathcal{V}_1} ||\mathcal{T}^s - \mathcal{X}^s||_F^2 + \sum_{(s,u)\in\mathcal{V}_2} ||\mathcal{T}^s - \mathcal{X}^s||_F^2 - \sum_{(s,u)\in\mathcal{V}_1\cup\mathcal{V}_2} ||\mathcal{T}^s - \mathcal{X}^s||_F^2 - \sum_{(s,u)\in\mathcal{V}_1\cap\mathcal{V}_2} ||\mathcal{T}^s - \mathcal{X}^s||_F^2) \\
&\quad + \mu_\ell/|\mathcal{S}| \sum_{k=1}^{n} (\sum_{(s,u)\in\mathcal{V}_1} \alpha_k^s ||\mathbf{A}_k^s||_F^2 + \sum_{(s,u)\in\mathcal{V}_2} \alpha_k^s ||\mathbf{A}_k^s||_F^2 - \sum_{(s,u)\in\mathcal{V}_1\cup\mathcal{V}_2} \alpha_k^s ||\mathbf{A}_k^s||_F^2 - \sum_{(s,u)\in\mathcal{V}_1\cap\mathcal{V}_2} \alpha_k^s ||\mathbf{A}_k^s||_F^2) \\
&\triangleq \mu_\ell/|\mathcal{S}|(f_1 + f_2)
\end{aligned} \tag{6}$$

Theorem 1 [1]. (***the partitioning property of CP decomposition***) *For a given streaming tensor sequence* $\{\mathcal{X}^{(t)}\}_{t=1}^T$*, if it could be approximated by* $[\![\mathbf{A}_1, \cdots, \mathbf{A}_N]\!]$*, and its sub-tensor* $\{\mathcal{X}^{(t)}\}_{t=1}^{T-1}$ *could be approximated by* $[\![\tilde{\mathbf{A}}_1, \cdots, \tilde{\mathbf{A}}_N]\!]$*, then* $\{\tilde{\mathbf{A}}_n\}_{n=1}^N$ *are the sub-matrix of* $\{\mathbf{A}_n\}_{n=1}^N$*, respectively.*

An illustration of the above property (in a three-way case) is demonstrated in Fig. 1. A given T-MUST tensor can be reconstructed into a corresponding multi-aspect streaming tensor described in [2], [3] by zero-padding (the green part).

For f_1, according to the definition of Frobenius norm, i.e., the square root of the sum of the absolute squares of tensor's elements, it can be seen that $f_1 = 0$. According to the partitioning property of CP decomposition, we can analyze the property of f_2 from two perspectives:

(1) $\mathcal{V}_1 \cup \mathcal{V}_2 = \emptyset$. In this case, we can deduce that $f_2 = 0$ according to the partitioning property of CP decomposition. Therefore, $J_\ell^l - J_\ell^r = \mu_\ell/|\mathcal{S}|(f_1 + f_2) = 0$.

(2) $\mathcal{V}_1 \cup \mathcal{V}_2 \neq \emptyset$. In this case, according to the partitioning property of CP decomposition, we can deduce that $f_2 \leq 0$. Therefore, $J_\ell^l - J_\ell^r = \mu_\ell/|\mathcal{S}|(f_1 + f_2) \leq 0$.

In conclusion, $J_\ell^l - J_\ell^r \leq 0$, that is, for $\forall \mathcal{V}_1, \mathcal{V}_2 \subseteq \mathcal{G}$, we have $J_\ell(\mathcal{V}_1) + J_\ell(\mathcal{V}_2) \leq J_\ell(\mathcal{V}_1 \cup \mathcal{V}_2) + J_\ell(\mathcal{V}_1 \cap \mathcal{V}_2)$, which means the first component J_ℓ is supermodular.

In addition, for any adding new element $v = (s_v, u_v)$ to $\mathcal{V}$ which means $z_{s_v u_v} = 1$, it incurs the constant marginal increment $2\mu_t \max_{(s,u)\in\mathcal{V}}\{\tau_{su}\}$ referring to the definition in Eq. (1)). Naturally, $\tilde{J}_t(\mathcal{V})$ is a linear increasing function which can be regarded as a monotone supermodular function as well.

Consequently, $\tilde{J}(\mathcal{V}) = \tilde{J}_t(\mathcal{V}) + J_\ell(\mathcal{V})$ is supermodular. ■

III. Proof of Lemma 1

Proof A3: When find the gap between $\tilde{J}(\mathcal{V})$ and $J(\mathcal{V})$, we only need compare $\tilde{J}_t(\mathcal{V})$ with $J_t(\mathcal{V})$. Without loss of generality, we present the function curve of both $\tilde{J}_t(\mathcal{V})$ and $J_t(\mathcal{V})$ in Fig. 2. Recall the expression for $J_t(\mathcal{V})$ and $\tilde{j}_t(\mathcal{V})$, we can easily derive $\tilde{J}_t(\mathcal{V}) \geq J_2(\mathcal{V})$ for any input $\mathcal{V}$. This is because $\tilde{J}_t(\mathcal{V})$ always has the maximal computation and model transfer latency ξ_{max} in the first component, and has the maximal communication latency $2\mu_t \max\limits_{(s,u)\in\mathcal{V}}\{\tau_{su}\}|\mathcal{V}|$ for any input $\mathcal{V}$ in the second component. Thus, we can always have $\tilde{J}_t(\mathcal{V}) \geq J_t(\mathcal{V})$ and $\tilde{J}_t(\mathcal{V})$ has a faster increasing speed than $J_t(\mathcal{V})$.

$\xi_{max} = \mu_t \max\limits_{s\in\mathcal{S}} \max\limits_{u\in\mathcal{U}}\{\gamma(\epsilon) t_{train}^s)\}$ When $\mathcal{V} = \emptyset$, that is, $|\mathcal{V}| = 0$, the difference between $\tilde{J}_t(\mathcal{V})$ and $J_t(\mathcal{V})$ is

$$\begin{aligned}
\Delta_{\mathcal{V}} &= \tilde{J}_t(\mathcal{V})_{|\mathcal{V}|=0} - J_t(\mathcal{V})_{|\mathcal{V}|=0} \\
&= \mu_t \max_{s\in\mathcal{S}} \max_{u\in\mathcal{U}}\{\gamma(\epsilon) t_{train}^s)\} = \xi_{max}.
\end{aligned}$$

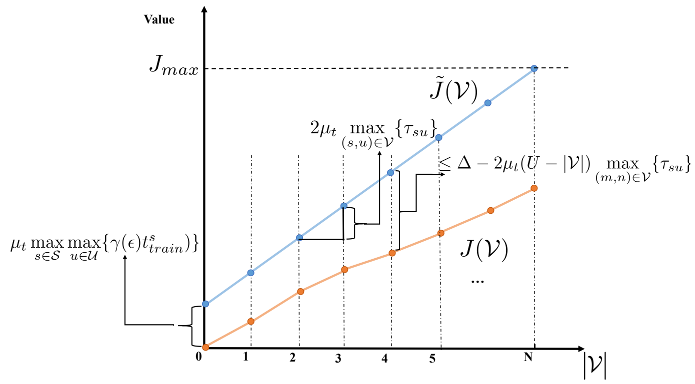

Fig. 2. Gap Analysis Between $J(\mathcal{V})$ and $\tilde{J}(\mathcal{V})$.

When $\mathcal{V}$ is not an empty set that $|\mathcal{V}| > 0, \mathcal{V} \in B(\mathcal{A})$, we have $J_t(\mathcal{V})_{|\mathcal{V}|=N} \geq \ \xi_{min} + 2\mu_t \min_{(s,u)\in\mathcal{V}}\{\tau_{su}\}$ and

$$\begin{aligned}
\Delta_{\mathcal{V}} \quad &= \tilde{J}_t(\mathcal{V}) - J_t(\mathcal{V}) \\
&\leq \xi_{max} + 2\mu_t \max_{(s,u)\in\mathcal{V}}\{\tau_{su}\}|\mathcal{V}| - (\xi_{min} + 2\mu_t \min_{(s,u)\in\mathcal{V}}\{\tau_{su}\} \\
&= \xi_{max} - \xi_{min} + 2\mu_t(|\mathcal{V}| \max_{(s,u)\in\mathcal{V}}\{\tau_{su}\} - \min_{(s,u)\in\mathcal{V}}\{\tau_{su}\}) \\
&= \Delta - 2\mu_t(U - |\mathcal{V}|) \max_{(m,n)\in\mathcal{V}}\{\tau_{su}\}.
\end{aligned} \tag{7}$$

Obviously, when $|\mathcal{V}| = U$, the different between $\tilde{J}_t(\mathcal{V})$ and $J_t(\mathcal{V})$ is bounded by Δ. Therefore, we can achieve $\tilde{J}(\mathcal{V}) \leq J(\mathcal{V}) + \Delta$ for any $\mathcal{V} \in B(\mathcal{A})$. ■

IV. The Edge Association Strategy of NTC-FL and Baselines on the Two Datasets

Table I and Table II show the edge association strategy of NTC-FL and baselines on Abilene and TaxiBJ-14, respectively. All UEs of the centralized baselines are associated with the default parameter server (*i.e.*, '#2' edge server).

TABLE I
The UE-Server Association Strategy on Abilene.

UEs	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
NTC-FL	#3	#2	#1	#2	#2	#1	#2	#3	#3	#3	#2	#1
NTC-FL/R	#1	#1	#1	#1	#1	#2	#2	#2	#3	#1	#1	#2
CP-WOPT	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2
SPC	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2
CoSTCo	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2
NTC	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2

TABLE II
The UE-Server Association Strategy on TaxiBJ-14.

UEs	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16
NTC-FL	#3	#7	#4	#8	#7	#4	#7	#6	#5	#2	#8	#2	#1	#6	#7	#5
NTC-FL/R	#2	#2	#2	#2	#2	#3	#3	#4	#6	#2	#1	#4	#3	#7	#5	#1
CP-WOPT	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2
SPC	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2
CoSTCo	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2
NTC	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2
UEs	#17	#18	#19	#20	#21	#22	#23	#24	#25	#26	#27	#28	#29	#30	#31	#32
NTC-FL	#8	#3	#1	#2	#6	#6	#4	#1	#4	#8	#3	#5	#5	#7	#3	#5
NTC-FL/R	#6	#1	#6	#2	#5	#2	#1	#1	#2	#8	#1	#8	#2	#8	#3	#8
CP-WOPT	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2
SPC	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2
CoSTCo	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2
NTC	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2	#2

REFERENCES

[1] C. Liu, T. Wu, Z. Li, T. Ma, and J. Huang, “Robust online tensor completion for iot streaming data recovery,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2022. [Online]. Available: https://doi.org/10.1109/TNNLS.2022.3165076

[2] M. Najafi, L. He, and P. S. Yu, “Outlier-robust multi-aspect streaming tensor completion and factorization,” in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, 2019, pp. 3187–3194.

[3] Q. Song, X. Huang, H. Ge, J. Caverlee, and X. Hu, “Multi-aspect streaming tensor completion,” in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, August 13 - 17, 2017*, 2017, pp. 435–443.