On scheduling algorithms for minimizing the energy consumption and execution time of Federated Learning

Laércio Lima Pilla (he/him)
laercio.lima-pilla@labri.fr
Agenda

- What is Federated Learning?
- Why do we care energy or carbon?
- What has been done by others?
- What have I been doing for performance?
- What have I been doing for energy?
- Where do we go from here?
What is Federated Learning?

**distributed learning + local data is never shared**

Communication-Efficient Learning of Deep Networks from Decentralized Data

![Diagram showing the process of Federated Learning with devices and a central FL server.](image-url)
What is Federated Learning?
distributed learning + local data is never shared
What is Federated Learning?

distributed learning + local data is never shared

Applications: next-word prediction, on-device item ranking, cyberattack detection, medical applications
Why do we care energy or carbon?

energy costs and emissions of ML

battery or energy available

convincing people to participate in training
Why do we care energy or carbon? energy costs and emissions of ML battery or energy available convincing people to participate in training
What has been done by others?

SmartPC: Hierarchical Pace Control in Real-Time Federated Learning System

Li Li*1, Haoyi Xiong*2,**, Jun Wang3, Cheng-Zhong Xu1,4, Zhishan Guo5,**,
1Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen China
2Big Data Lab (BDL) and PaddlePaddle, Baidu, Inc., Beijing, China
3Department of Electrical and Computer Engineering, McGill University, Montréal, Canada
4Department of Computer and Information Science, University of Macau, Macau, China
5Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL

[RTSS’19]

• Define a virtual deadline
• Employ DVFS on devices

Completion Time Model. With the received hardware configuration information and the size of local training data, the time required by device $i$ to complete the local training process can be modeled [6] as:

$$t_i = \frac{c_i D_i}{f_i}$$  \hspace{1cm} (1)
What has been done by others?

Energy Efficient Federated Learning Over Wireless Communication Networks

Zhaohui Yang\textsuperscript{d}, Member, IEEE, Mingzhe Chen\textsuperscript{d}, Member, IEEE, Walid Saad\textsuperscript{d}, Fellow, IEEE, Choong Seon Hong\textsuperscript{d}, Senior Member, IEEE, and Mohammad Shikh-Bahaei\textsuperscript{d}, Senior Member, IEEE


• **Minimize energy** under a latency constraint
• Control the transmission time, bandwidth allocation, clock frequency, transmission power, and learning accuracy

1) Local Computation: Let $f_k$ be the computation capacity of user $k$, which is measured by the number of CPU cycles per second. The computation time at user $k$ needed for data processing is

$$
\tau_k = \frac{I_k C_k D_k}{f_k}, \quad \forall k \in \mathcal{K},
$$

where $C_k$ (cycles/sample) is the number of CPU cycles required for computing one sample data at user $k$ and $I_k$ is the number of local iterations at user $k$. According to Lemma 1 in [35], the energy consumption for computing a total number of $C_k D_k$ CPU cycles at user $k$ is

$$
E_{k1}^C = \kappa C_k D_k f_k^2.
$$
What have I been doing for performance?

Problem: given $T$ tasks and $n$ resources with different cost functions $C$, find an assignment that minimizes the maximal cost $C_{\text{max}}$.

Assumption: all cost functions are monotonically increasing and known.

![Diagram showing cost functions $C_1$ and $C_3$ versus tasks]
Laércio Lima Pilla - On scheduling algorithms for FL

What have I been doing for performance?

**Optimal Task Assignment for Heterogeneous Federated Learning Devices**

Laércio Lima Pilla

[IPDPS’21]

- Minimize training round duration
- Control the amount of work on each device
- Assuming increasing costs
- Optimal solution with time in $O(n + T \log n)$

**Problem**

Find a task assignment $A_i \in N$ to each resource $i \in R$ that minimizes the makespan $C_{\text{max}}$ while assigning all tasks among the resources and respecting their lower and upper limits.

\[
C_{\text{max}} = \max_{i \in R} C_i(A_i)
\]

\[
\sum_{i \in R} A_i = T
\]

\[
L_i \leq A_i \leq U_i, \ \forall i \in R
\]
What have I been doing for performance?

Optimal Task Assignment for Heterogeneous Federated Learning Devices

Laercio Lima Pilla

Algorithm 1: OOLAR

Data: Tasks $T$, Resources $R$, Cost functions $C_i(\cdot)$, Lower and Upper limits $L_i$ and $U_i$ ($i \in R$)

Result: Assignment of tasks to resources $A_i$ ($i \in R$)

1. $h \leftarrow \text{min-heap}(\cdot)$ ▶ Heap sorted by cost
2. for $i \in R$ do
3.   $A_i \leftarrow L_i$ ▶ Resources start at their lower limit
4.   if $A_i < U_i$ then
5.     $h.push(C_i(A_i + 1), i)$
6.   end
7. end
8. for $t$ from $1 + 1$ to $T$ do
9.   $(c, j) \leftarrow h.pop()$
10.   $A_j \leftarrow A_j + 1$ ▶ Assigns $t$ to $j$
11. if $A_j < U_j$ then
12.     $h.push(C_j(A_j + 1), j)$
13. end
14. end
What have I been doing for energy?

Problem: given T tasks and n resources with different cost functions C, find an assignment that minimizes the total cost $\sum C$.

Assumption: all cost functions are known
What have I been doing for energy?

Scheduling Algorithms for Federated Learning With Minimal Energy Consumption

Laércio Lima Pilla

[TPDS 2023]

- Minimize total [energy] cost
- Control the amount of work on each device
- Optimal solution with time in $O(T^2n)$
  - *Multiple-Choice Minimum-Cost Maximum Knapsack Packing Problem* - (MC)$^2$MKP
- Faster solutions for cases with specific marginal cost behaviors

**Problem**

Find a task assignment $A_i \in \mathbb{N}$ to each resource $i \in R$ that minimizes the total cost $\Sigma C$ while assigning all tasks among the resources and respecting their lower and upper limits.

$$\Sigma C = \sum_{i \in R} C_i(A_i)$$

$$\sum_{i \in R} A_i = T$$

$$L_i \leq A_i \leq U_i, \ \forall i \in R$$
What have I been doing for energy?

Scheduling Algorithms for Federated Learning With Minimal Energy Consumption

Laércio Lima Pilla

(MC)²MKP formulation

\[
\text{minimize}_{x,y} \sum_{i=1}^{n} \sum_{j \in N_i} c_{ij} x_{ij} - y \sum_{i=1}^{n} \sum_{j \in N_i} c_{ij}, \quad (2a)
\]

subject to \[
\sum_{i=1}^{n} \sum_{j \in N_i} w_{ij} x_{ij} \leq T, \quad (2b)
\]

\[
\sum_{i=1}^{n} \sum_{j \in N_i} w_{ij} x_{ij} \geq y, \quad (2c)
\]

\[
\sum_{j \in N_i} x_{ij} = 1, \quad i = 1, \ldots, n, \quad (2d)
\]

\[
x_{ij} \in \{0, 1\}, \quad i = 1, \ldots, n, \quad j \in N_i, \quad (2e)
\]

\[
y \in [0, T] \quad (2f)
\]
What have I been doing for energy?

Scheduling Algorithms for Federated Learning With Minimal Energy Consumption

Laércio Lima Pilla

(MC)$^2$MKP formulation

General dynamic programming idea

1. Build all possible solutions for the first device
2. Build all possible solutions considering the best solutions found for i devices (1 ≤ i ≤ n-1)
What have I been doing for energy?

Other scenarios with **monotonically increasing costs**

1. **Increasing** marginal costs (**MarIn**): similar to OLAR
2. **Constant** marginal costs (**MarCo**): assign as many tasks as possible to the devices with the smallest constant marginal costs
3. **Decreasing** marginal costs without upper limits (**MarDecUn**): assign everything to the device with the smallest total cost
4. **Decreasing** marginal costs (**MarDec**): split resources between those with and without upper limits, test all possible Minimum-Cost Maximum Knapsack Packings

| Solutions With the Smallest Complexity for the Variations of Our Scheduling Problem. We Assume $T > n$ |
|---|---|---|---|
| | Arbitrary Costs | Increasing | Marginal Costs |
| | Constant | Decreasing |
| Without upper limits | $(MC)^2\text{MKP} - O(T^2n)$ | $\Theta(T \log n)$ | $\Theta(n)$ | $\Theta(n)$ |
| With upper limits | $(MC)^2\text{MKP} - O(T^2n)$ | $\Theta(T \log n)$ | $\Theta(n \log n)$ | $\Theta(Tn^2)$ |

Laércio Lima Pilla - On scheduling algorithms for FL
What have I been doing for energy?

**Advantages of this kind of approach**

- Given the required information, we can find optimal solutions
- It should be easier to control how much work to give to a resource than to control other aspects of the resources

**Limitations**

- We require more information from the resources
- Optimizing for one objective can lead to worse outcomes for other objectives
Where do we go from here?

We take FL and optimize its performance or energy consumption by controlling how much work each device should do.

- **Does this apply elsewhere?** (workload distribution problems)
- Can we get energy or carbon-equivalent emissions information?
  - Should we? (privacy issues)
  - Can we trust this information?
- How does this affect convergence/training time? (other metrics)
- Can we optimize for energy + something? How to combine?
  - **How far from the best performance** (%) do I accept to be if it improves my energy consumption?
  - Idem, but switching performance and energy
- Can we adapt dynamically?

That’s all folks!