FL & CV

Presenter: Ziqi Zhou

Mentor: Dr.Hao Wang

What is federated learning?

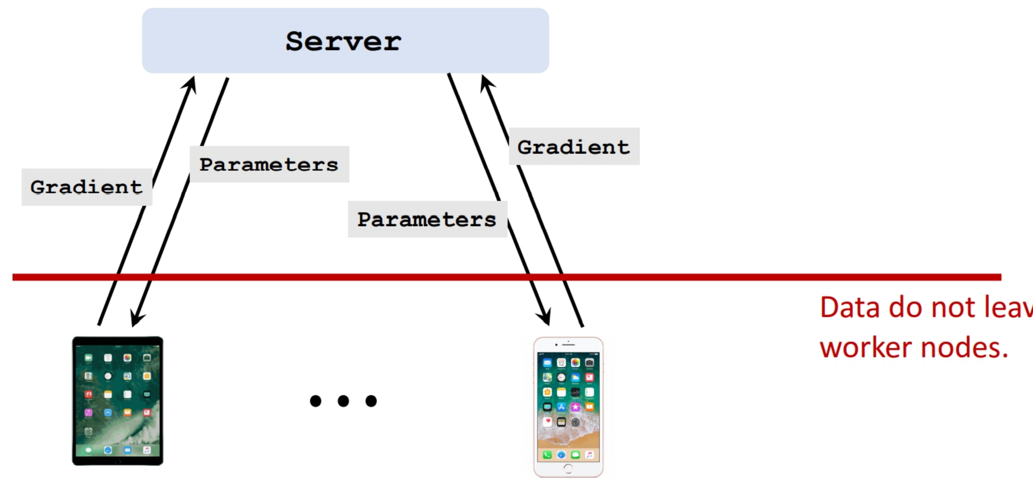
Federated learning is a kind of distributed learning.

How does federated learning differ from traditional distributed learning?

- 1. Users have control over their device and data.
- 2. Worker nodes are unstable.
- 3. Communication cost is higher than computation cost.
- 4. Data stored on worker nodes are not IID.
- 5. The amount of data is severely imbalanced.

Motivation

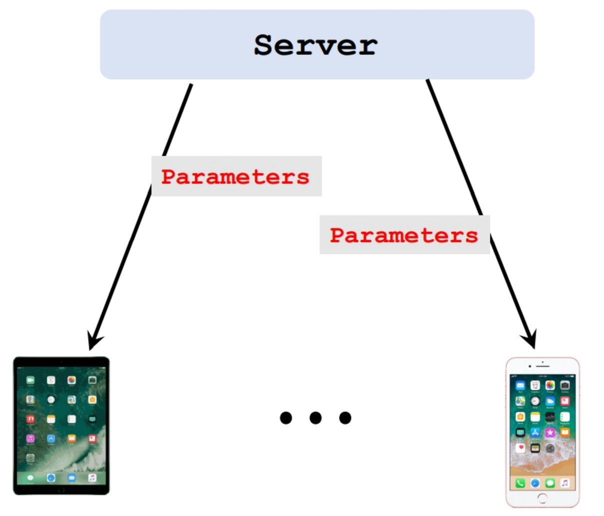
When we are doing CV tasks like facial expressions recognition, requiring large amount of data from clients, privacy becomes an unavoidable problem. So it's quite necessary for us to figure out a way that doesn't require data sharing. And federated learning would be a direction worth trying.



Data do not leave worker nodes.

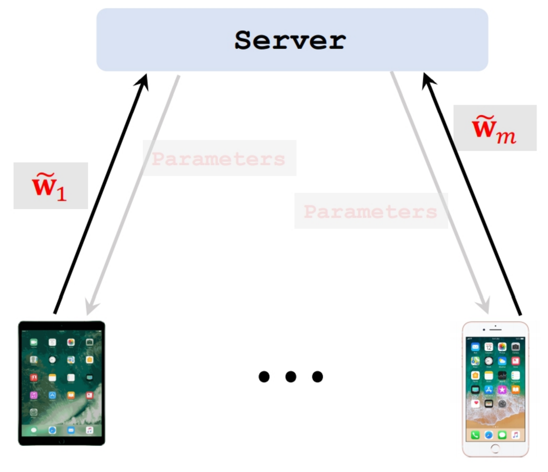
Computation vs. Communication

Federated Averaging Algorithm



The i-th worker performs:

1. Receiving model parameters $\mathbf{w}$ from the server.
2. Repeating the followings:
   a) Using $\mathbf{w}$ and its local data to compute gradient $\mathbf{g}$.
   b) Local update: $\mathbf{w} \leftarrow \mathbf{w} - \alpha \cdot \mathbf{g}$.
3. Sending $\tilde{\mathbf{w}}_i = \mathbf{w}$ to the server.



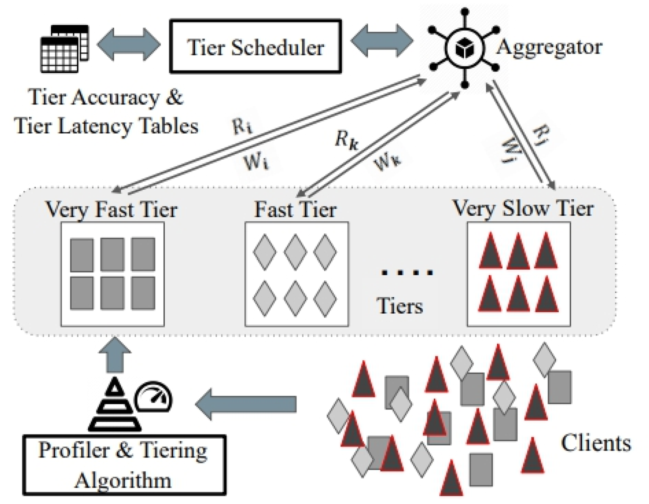
The server performs:

1. Receiving $\tilde{\mathbf{w}}_1, \cdots, \tilde{\mathbf{w}}_m$ from all the m workers.
2. Updating model parameters:

$$\mathbf{w} \leftarrow \frac{1}{m}(\tilde{\mathbf{w}}_1 + \cdots + \tilde{\mathbf{w}}_m).$$

Resource & data heterogeneity

TiFL: A Tier-based Federated Learning System

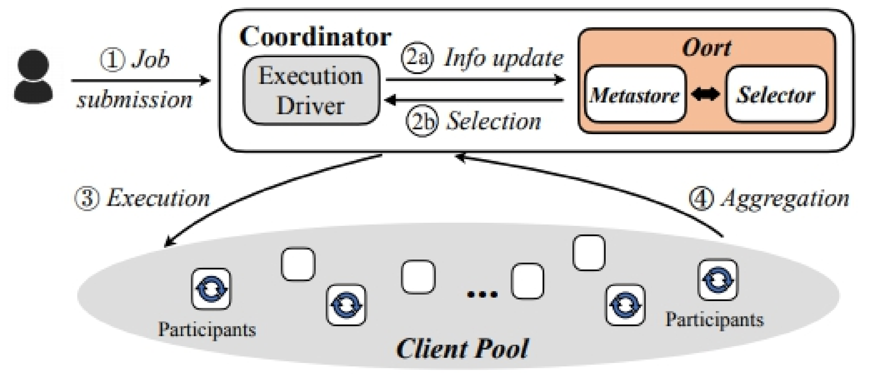


Algorithm 2 Adaptive Tier Selection Algorithm. $Credits_t$: the credits of Tier t, I: the interval of changing probabilities, $TestData_t$: evaluation dataset specific to that tier t, A_t^r: test accuracy of tier t at round r, τ: set of Tiers.

Aggregator: initialize weight w_0, $currentTier = 1$, $TestData_t$, $Credits_t$, equal probability with $\frac{1}{T}$, for each tier t.
for each round $r = 0$ **to** $N-1$ **do**
  if $r\%I == 0$ and r ≥ I **then**
    if $A^r_{currentTier} \leq A^{r-I}_{currentTier}$ **then**
      $NewProbs = ChangeProbs(A_1^r, A_2^r ... A_T^r)$
    end if
  end if
  while $True$ **do**
    $currentTier$ = (select one tier from T tiers with $NewProbs$)
    if $Credits_{currentTier} > 0$ **then**
      $Credits_{currentTier} = Credits_{currentTier} - 1$
      break
    end if
  end while
  C_r = (random set of $|C|$ clients from $currentTier$)
  $Credits_{currentTier} -= 1$
  for each client c ∈ C_r **in parallel do**
    $w_r^c = TrainClient(c)$
    s_c = (training size of c)
  end for
  $w_r = \sum_{c=1}^{|C|} w_{r+1}^c * \frac{s_c}{\sum_{c=1}^{|C|} s_c}$
  for each t in τ **do**
    $A_t^r = Eval(w_r, TestData_t)$
  end for
end for

Resource & data heterogeneity

Oort: Informed Participant Selection for Scalable Federated Learning



Input: Client set $\mathbb{C}$, sample size K, exploitation factor ε, pacer step Δ, step window W, penalty α
Output: Participant set $\mathbb{P}$

/* Initialize global variables. */
$\mathbb{E} \leftarrow \emptyset$; $\mathbb{U} \leftarrow \emptyset$ ▷ Explored clients and statistical utility.
$\mathbb{L} \leftarrow \emptyset$; $\mathbb{D} \leftarrow \emptyset$ ▷ Last involved round and duration.
$R \leftarrow 0$; $T \leftarrow \Delta$ ▷ Round counter and preferred round duration.

/* Participant selection for each round. */
Function SelectParticipant ($\mathbb{C}, K, \varepsilon, T, \alpha$)
  $Util \leftarrow \emptyset$; $R \leftarrow R+1$
  /* Update and clip the feedback; blacklist outliers. */
  UpdateWithFeedback($\mathbb{E}, \mathbb{U}, \mathbb{L}, \mathbb{D}$)
  /* Pacer: Relaxes global system preference T if the statistical utility achieved decreases in last W rounds. */
  if $\sum \mathbb{U}(R-2W : R-W) > \sum \mathbb{U}(R-W : R)$ **then**
    $T \leftarrow T + \Delta$
  /* Exploitation #1: Calculate client utility. */
  for *client* $i \in \mathbb{E}$ **do**
    $Util(i) \leftarrow \mathbb{U}(i) + \sqrt{\frac{0.1 \log R}{\mathbb{L}(i)}}$ ▷ Temporal uncertainty.
    if $T < \mathbb{D}(i)$ **then** ▷ Global system utility.
      $Util(i) \leftarrow Util(i) \times (\frac{T}{\mathbb{D}(i)})^{\alpha}$
  /* Exploitation #2: admit clients with greater than $c\%$ of cut-off utility; then sample $(1-\varepsilon)K$ clients by utility. */
  $Util \leftarrow$ SortAsc(Util)
  $\mathbb{W} \leftarrow$ CutOffUtil($\mathbb{E}$, $c \times Util((1-\varepsilon) \times K)$)
  $\mathbb{P} \leftarrow$ SampleByUtil($\mathbb{W}$, $Util$, $(1-\varepsilon) \times K$)
  /* Exploration: sample unexplored clients by speed. */
  $\mathbb{P} \leftarrow \mathbb{P} \cup$ SampleBySpeed($\mathbb{C} - \mathbb{E}$, $\varepsilon \times K$)
  return $\mathbb{P}$

Challenges

Data form heterogeneity

- people usually mix all photos from their lives altogether, in which there might be selfies, scenic photos, vlogs, screenshots, etc.
- images with different number of faces also need to be processed differently.

Gender/race bias

- users tend to save photos of their own on their devices.
- the genders and races of people in the photos are more likely to be the same.

Emotion bias

- people mostly record their happy moments in their mobiles, while few would save images with negative emotions on purpose.

References

[1]H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," arXiv:1602.05629 [cs], Feb. 2016, Accessed: Jun. 07, 2019.

[2]Z. Chai et al., "TiFL: A Tier-based Federated Learning System," in Proceedings of the 29th International Symposium on High-Performance Parallel and Distributed Computing, Stockholm Sweden, Jun. 2020, pp. 125–136.

[3]F. Lai, X. Zhu, H. V. Madhyastha, and M. Chowdhury, "Oort: Informed Participant Selection for Scalable Federated Learning," arXiv:2010.06081 [cs]

[4]Kovac, J., Peer, P. and Solina, F. (2003) *Human Skin Colour Clustering for Face Detection.* International Conference on Computer as a Tool EUROCON, 2, 144-148.

[5]Cao, X.Y., Liu, H.F., 2011. *A Skin Detection Algorithm Based on Bayes Decision in the YCbCr Color Space*. AMM 121–126, 672–676.

[6]M. Duan, D. Liu, X. Chen, R. Liu, Y. Tan and L. Liang, "Self-Balancing Federated Learning With Global Imbalanced Data in Mobile Systems," in IEEE Transactions on Parallel and Distributed Systems, vol. 32, no. 1, pp. 59-71, 1 Jan. 2021, doi: 10.1109/TPDS.2020.3009406.

Acknowledgement

This work was funded by an unrestricted gift from Google. I would like to thank UTRGV for hosting this program.