

Seizing Critical Learning Periods in Federated Learning

Gang Yan[†], Hao Wang[‡] and Jian Li[†]

[†]Department of Electrical and Computer Engineering, SUNY-Binghamton University

[‡]Division of Computer Science and Engineering, Louisiana State University



Introduction

Critical learning periods determine the final quality of a deep neural network (DNN) in traditional ML. During a critical period, deficits such as low quality or quantity of training data will cause irreversible model degradation.

In this work, our main contributions are:

1. Discover that critical learning periods consistently exist in Federated Learning (FL).
2. Systematically explore the impacts of critical learning periods for FL under a wide range of FL hyperparameters.
3. Propose a new notion dubbed Federated Fisher Information Matrix and analyze the phenomenon of critical learning periods through trace of FedFIM.

Notations and Hypothesis

Notations are:

1. \mathbf{w} : denotes model parameters, training round is $t = 1, \dots, T$.
2. \mathcal{N} : set of total clients, the selected subset of clients is $\mathcal{N}_t \subset \mathcal{N}$.
3. \mathcal{D}_j : local dataset of client j , entire training dataset is $\mathcal{D} = \cup_{j \in \mathcal{N}} \mathcal{D}_j$.
4. η_j : learning rate of client j .

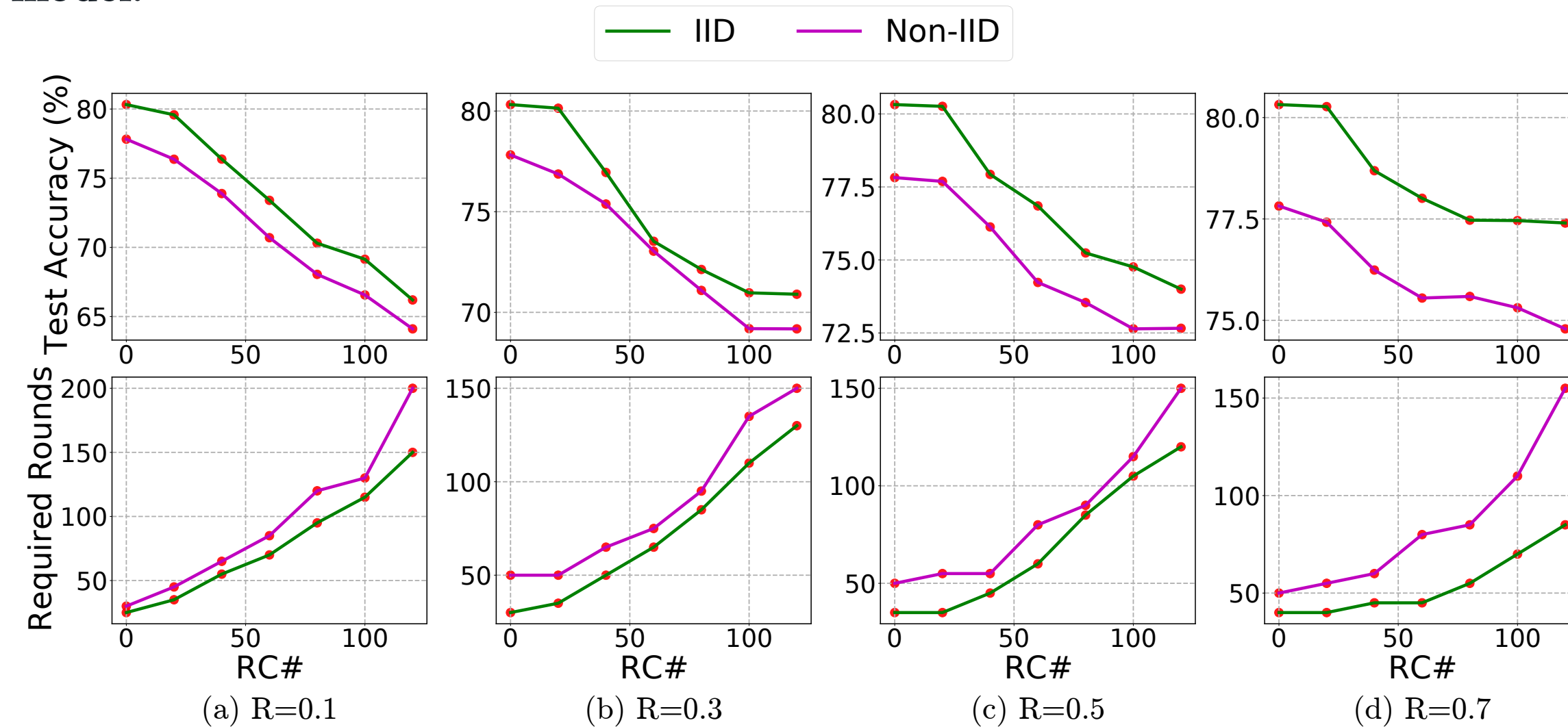
To find critical periods in Federated Learning, we hypothesi that:

1. A model with loss function $\ell(\mathbf{x}; \mathbf{w})$, where ℓ reaches a minimum loss ℓ_{loss} with a test accuracy ℓ_{acc} .
2. Only a subset of the local training data $\mathcal{D}'_j \subset \mathcal{D}_j, \forall j \in \mathcal{N}$ used to train model in the first M rounds and then using entire training dataset \mathcal{D} afterwards.
3. ℓ reaches a minimum loss of $\ell'_{\text{loss}}(M)$ with a test accuracy of $\ell'_{\text{acc}}(M)$.
4. There exist M_1 and M_2 such that $\ell'_{\text{acc}}(M_1) \geq \ell'_{\text{acc}}(M_2)$ when $M_1 \leq M_2$.

* M is defined as Recover Round (RC#).

Existence of Critical Learning Periods

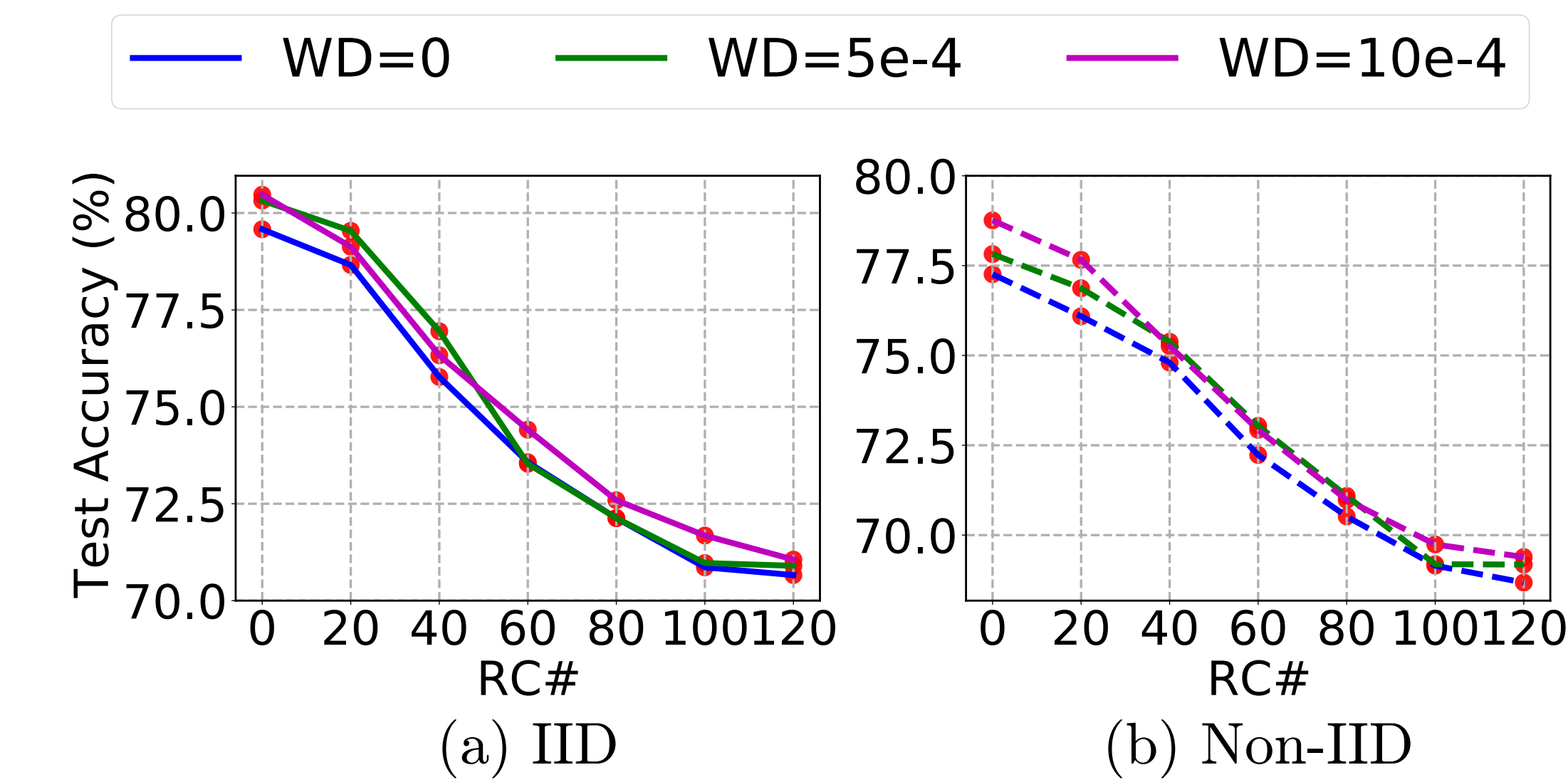
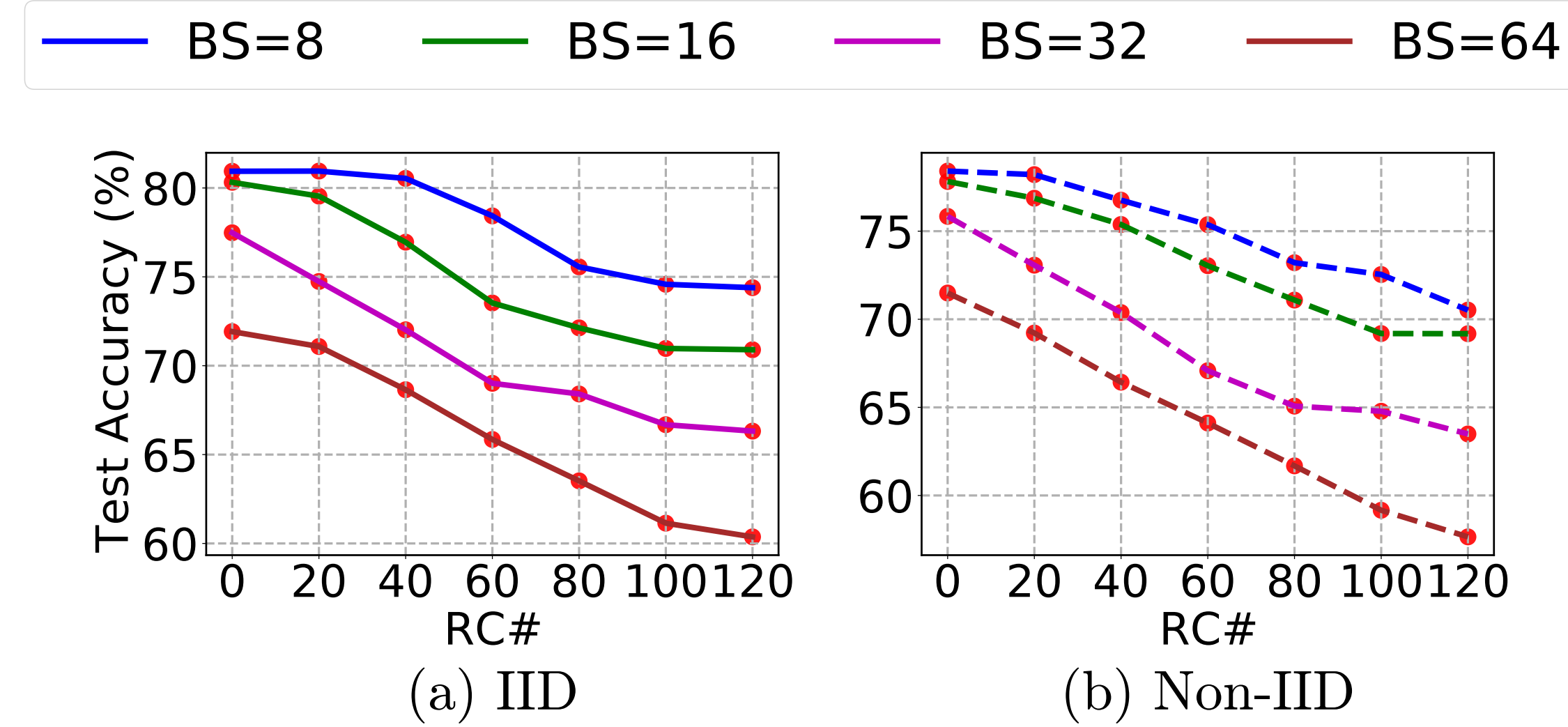
Using ResNet-18 and CIFAR-10, and R is percentage of local data to train model:



- Critical learning periods consistently exist in FL.
- More rounds to train model with partial data, worse the final accuracy.
- More rounds to train model with partial data, more rounds to achieve a predefined accuracy.

Hyperparameters

Similar results are also observed under different batch size, weight decay.



- Critical learning periods consistently exist regardless of the choice of hyperparameters.
- The phenomenon of critical learning periods in FL cannot be simply explained by the differences in hyperparameters.

Metrics

Three metrics used to explain critical learning periods:

- Fisher Information Matrix (FIM):

$$\mathbf{F}(\mathbf{w}) = \mathbb{E}_{\mathbf{x} \sim \mathcal{X}, \hat{y} \sim p_{\mathbf{w}}(y|\mathbf{x})} [g(\mathbf{x}, \hat{y})g(\mathbf{x}, \hat{y})^T], \quad (1)$$

where where $p_{\mathbf{w}}(y|\mathbf{x})$ is probabilistic model and $g(\mathbf{x}, \hat{y})$ is loss gradient

- Federated Fisher Information Matrix (FedFIM):

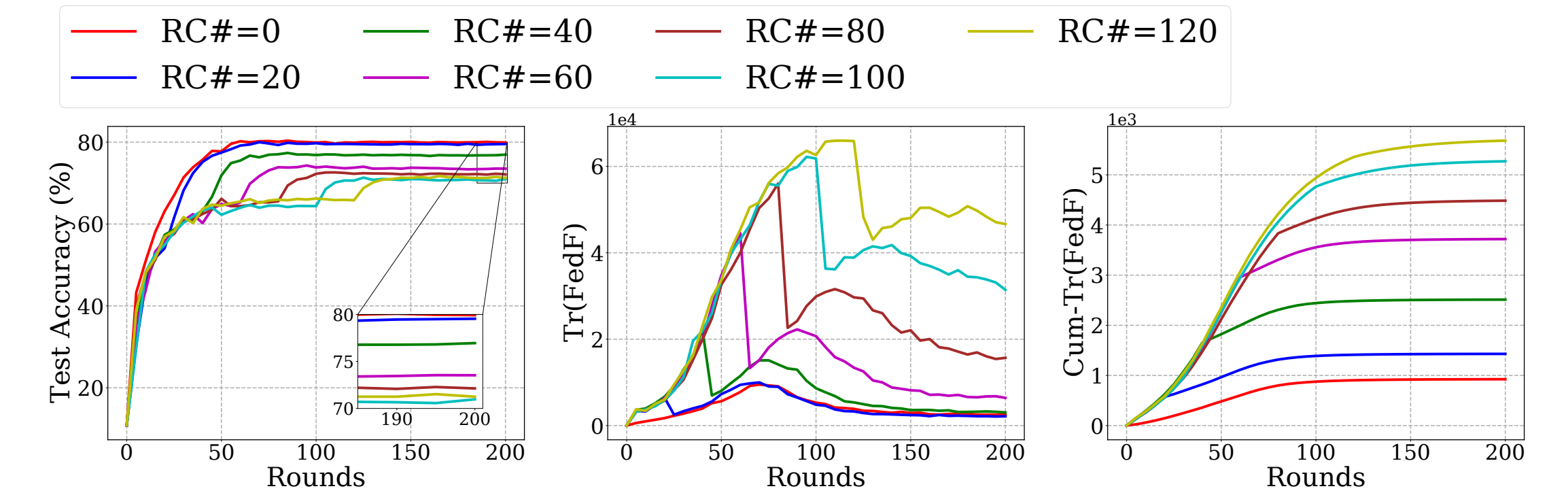
$$\text{FedF}(\mathbf{w}) = \sum_j \frac{|\mathcal{D}_j|}{|\mathcal{D}|} \mathbf{F}_j(\mathbf{w}), \quad (2)$$

- Weighted Cumulative Sum of the Trace of FedFIM (Cum-Tr(FedFIM))

$$\text{Cum-Tr}(\text{FedF})(k) = \sum_{i=0}^k \eta_i \text{Tr}(\text{FedF}_i), \quad (3)$$

Explanation

Using $R = 0.3$ as an illustrating example.

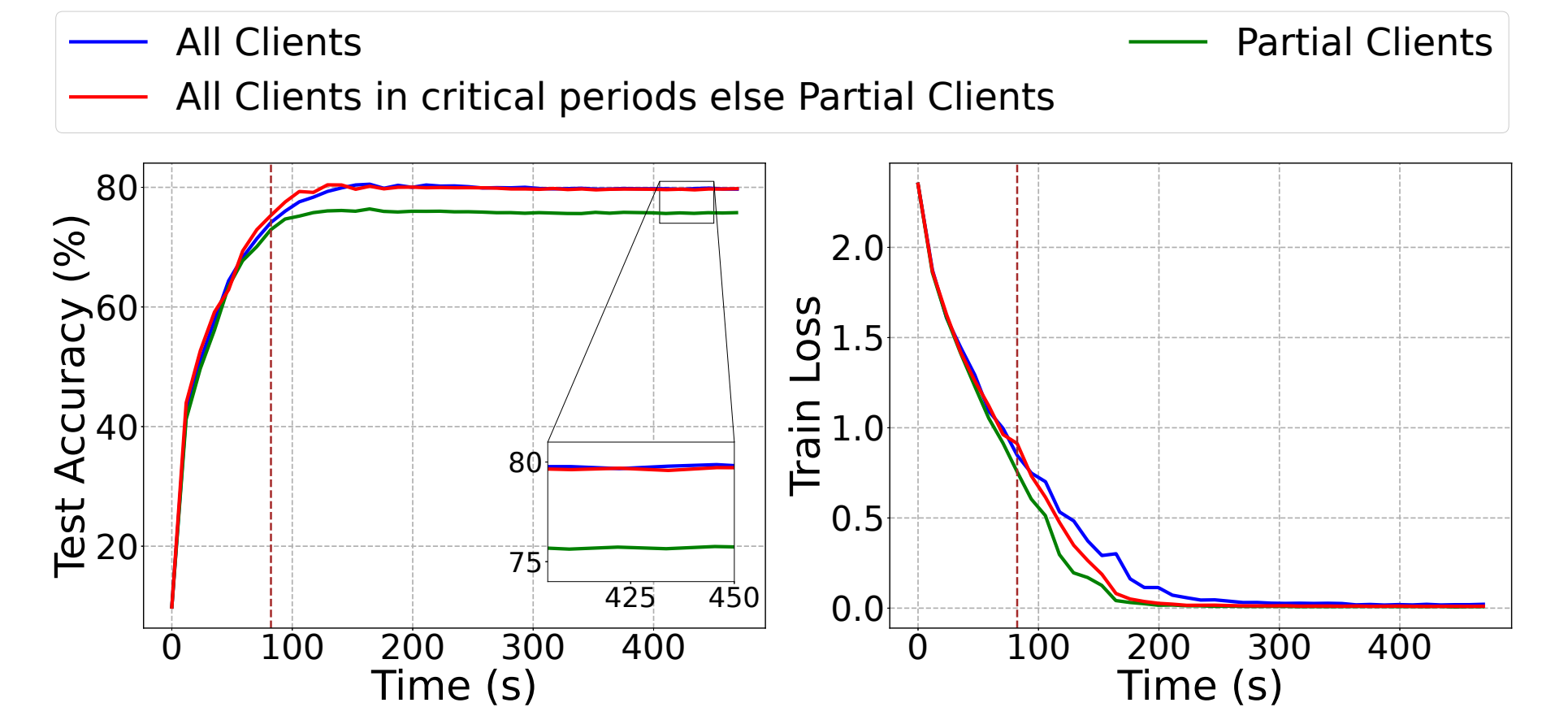


- Trace of FedFIM represents degree of whether local data is good enough.
- Sharp increase in trace in early phases, which coincides with accuracy.
- Additional data further boosts the test accuracy.

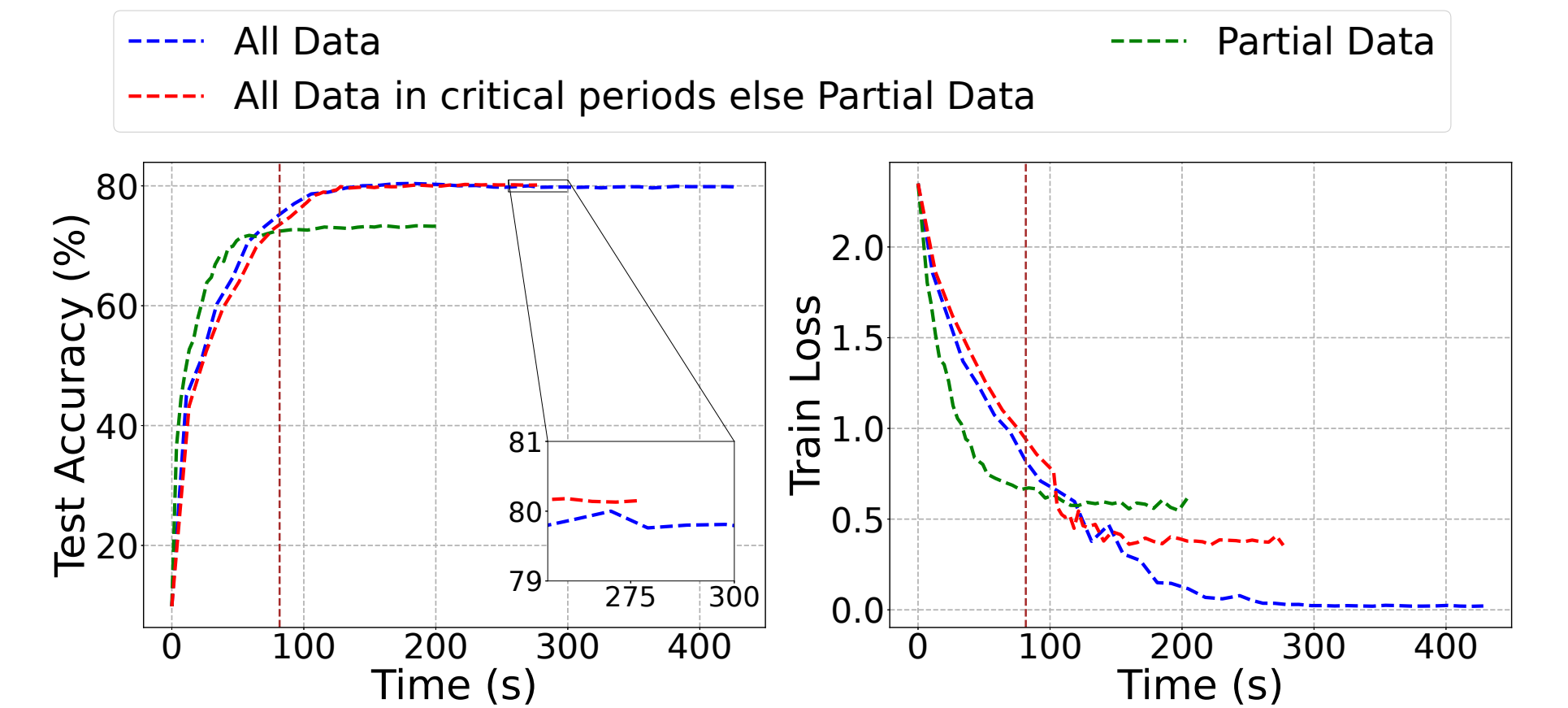
Seizing Critical Learning Periods

Use critical learning periods to boost training efficiency

- No need to involve all clients throughout the training process.



- No need to use the entire training data throughout the training process.



- Leveraging critical learning periods improves training efficiency.

Conclusion

Critical learning periods in FL improve training efficiency of global model.

- Show that critical learning periods consistently exist in FL.
- Propose a new metric called Federated Fisher Information Matrix.
- Suggest that the initial learning phase plays a critical role.