Horizontal Federated Learning Frameworks

AUTHORS

Márton Soos - 4913760 - m.soos@student.tudelft.nl

Responsible Professor: Dr. Kaitai Liang
Supervisor: Rui Wang

1 BACKGROUND

- Federated Learning is a type of distributed machine learning
- In HFL, data on each client has the same set of features
- Classic HFL algorithm is vulnerable to:
  - inferrence attacks
  - model/data poisoning
- These mitigated at the cost of performance and accuracy, using:
  - Differential Privacy
  - Homomorphic Encryption
  - Secure Multi-Party computation

2 OBJECTIVE

This literature study aims to answer the following questions:

1. How is Horizontal Federated Learning implemented?
2. What privacy and performance trade-offs are made when designing HFL frameworks?
3. How do HFL frameworks compare in terms of computational complexity, communication cost and privacy guarantees?

3 METHODOLOGY

7 Different HFL frameworks were studied. For each of them a theoretical analysis was performed to determine:

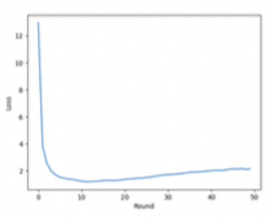
1. The time complexity
2. Communication cost
3. Security/privacy guarantees

For two of the frameworks experiments have been reproduced to assess the accuracy of their resulting models.

5 EXPERIMENTS

- Two of the frameworks have been simulated to reproduce experiments
- The convergence and accuracy of the final models was assessed
- Most of the results were successfully reproduced

Dataset	Original Accuracy	Reproduced Accuracy
Credit Card	81.09	78
Breast Cancer	95.62	95.91
Audit Data	97.42	97.4359



4 FRAMEWORK ANALYSIS

Peer-to-Peer MPC:

- Models are trained on each participant - O(T)
- Gradients are aggregated using the Additive Secret-Sharing MPC Protocol[1]:
  - Each participant generates C secret shares for their gradient, sends 1 to each of the other participants - O(C*S)
  - Secret shares are summed locally, then the sums are broadcasted and summed to obtain final results - O(C*S)

Two-Phase-MPC

- A committee of K participants is elected - O(C)
- Each participant generates K secret shares, sends them to committee- O(K*S)
- The committee aggregates the shares - O(C*S + K*S)
- The committee sends the result to each participant - O(C/K*S)

Framework	Time Complexity
Classic HFL	$O(R*(T+C*S))$
FederBoost	$O(M*logN*(logn+C))+M*n+M*C)$
GRAFFL	$O(C+T_{SuffiAE}+R*n*d+R*N*log(R*N))$
SplitFed	$O(R*(S_C+L_S+C*L_C))$
Fusion Learning	$O(M+T_L+C*(S_L+ng)+T_G)$
Peer-to-Peer MPC	$O(R*(T+C*S))$
Two-Phase MPC	$O(C+R*(T+K*S+C*S+C/K*S))$
FLOP	$O(R*(T+C+S_S))$
PFMLP	$O(R*(T+k^2*\alpha*S+C*S))$

Framework	Privacy-preserving measure	Ensures privacy against
Classic HFL	None	None
FederBoost	Simplified Masking Protocol	Honest-but-curious clients
GRAFFL	Using summaries of private data Not sharing gradients/weights	Fully-dishonest, curious (and colluding) server and/or clients
SplitFed	Differential Privacy	Fully-dishonest, curious (and colluding) server and/or clients
Fusion Learning	Not sharing gradients/weights	None
Peer-to-Peer MPC	Secure MPC	Honest-but-curious participants without collusion
Two-Phase MPC	Secure MPC	Honest-but-curious participants without collusion
FLOP	Only sharing part of model	Unknown
PFMLP	Homomorphic Encryption	Honest-but-curious server, keyserver and clients without collusion

6 CONCLUSION

- Multiple approaches exist to HFL each with their own advantages and disadvantages
-

TU Delft

RELATED LITERATURE
[1] Renuga Kanagavelu et al. "Two-Phase Multi-Party Computation Enabled Privacy-Preserving Federated Learning". In:2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID). 2020, pp. 410–419.DOI: 10 . 1109 /CC Grid49817.2020.00-52.