Distributed Federated Learning in Wireless Ad hoc Networks

Evangelia Fragkou *and* Dimitrios Katsaros

Department of Electrical and Computer Engineering,
University of Thessaly, Volos

Abstract

Federated deep learning has become a very popular deep learning methodology for performing distributed machine learning in environments where data sharing is not allowed due to privacy/security issues. However, in wireless ad hoc networks the traditional federated learning concept with a single coordinator (server) that performs the global averaging is not viable due to the excessive communication overhead. Therefore, a different architecture that is appropriate for the networking environment has been very recently proposed, that of decentralized (purely distributed or device-to-device) federated learning. Nevertheless, all the proposals present some deficiencies: a) they either are hybrid in nature (still assume the existence of a central server), b) they work on flat ad hoc networks, c) they select the clients' participation in a data-agnostic way. In this poster, we develop a purely distributed federated deep learning architecture based on appropriate network clustering, and on a data-aware client participation protocol in the federation. We evaluate the proposed methodology on real and synthetic data targeting a time-series prediction task, using a Gated Recurrent neural network (GRU). The obtained results confirm the merits of our approach.

Introduction & Motivation

Traditionally, ML is performed in a centralized manner where both data and processing are stored in a powerful server (or cloud) [1]. However, a lot of applications are based on distributedly interconnected devices, which generate private data that cannot be sent over the network to the "server". For such situations, Federated Learning (FL) has emerged a few years ago to address problems e.g. privacy concerns. In this architecture, nodes perform local neural network model training and send the model's parameters to the server for global averaging. Still, this architecture is not appropriate for wireless ad hoc environments. To be adjusted to them, FL schemes with either device-to-device communication and a coordinating server, or only device-to-device communication, have been proposed, however, they both do not fully exploit the properties of a wireless ad hoc environment such as that found in Internet of Things (IoT). Some very recent proposals, though, assumed a clustered ad hoc network in FL environments, however, even these works have not managed to address the peculiarities of these kind of network, since they do not pay attention to the actual clustering mechanism, which plays a key role on communication reduction.

Taking into consideration that the vast majority of the ad hoc network clustering schemes – if not all – are based on node IDs, it is very common for the cluster heads (CH), i.e., the nodes that will eventually perform the local averaging in the FL scheme, to be one-hop neighbors. This will negatively impact the communication efficiency of the scheme due to the interference since all the intra-cluster communication will be carried out by the cluster head. Moreover, almost all ad hoc clustering algorithms produce clusters with very small cardinality, thus local averaging/consensus is not very efficient with respect to the communication reduction. Additionally, the communication range restrictions (a range of a few hundred meters) results in the wireless nodes being frequently in very close proximity, thus producing very similar data, e.g., sensor measuring temperature. This situation turns almost all proposed schemes for client participation to the federation, e.g., based on the past workload completion capability, on secretary selection formulation as being not appropriate.

Based on the above shortcomings, in the present article we contribute mainly the following methods towards developing an appropriate federated deep learning algorithm for wireless ad hoc network environments [2]:

• We develop an ad hoc network clustering scheme appropriate for performing local and global aggregations that reduces wireless interference,

• We develop a client participation to federation scheme that is based on data-similarity to reduce communication during local aggregation,

• We develop a suite of federation algorithms based on variations of the basic methods, and evaluate them based on speed of convergence.

Clustering & data-similarity-based node participation (FEDLwa_SNP)

In order to reduce communication:

- We generate large clusters, using the max-min d-cluster formation protocol, which is based on dominating sets and is of O(n) complexity,
- Then, betweenness centrality measures are used in order to find the optimal CH (to the "center"),
- Data of adjacent network nodes (e.g.,sensors) may be correlated due to their geographical proximity. Therefore not all nodes are important to the same extent, regarding the training procedure and we can exclude them,
- We used Discrete Fourier Transform (DFT) to transform the time-series data from the time domain into the energy domain. Then *Parseval's* theorem [3] implies that the DFT preserves the Euclidean distance between two "signals",
- In practice, in each training round, data of each node pass through the Fast Fourier Transform function (FFT), providing a list of integers, representing the DFT coefficients of the data. Subsequently, we keep and transmit the first few integers to the CH to implement the comparison, since they are the most significant ones and can be both a small and effective sample.

Theorem 1. *Let X be the DFT if a sequence x. Then, the following holds*: $\sum_{i=0}^{n-1}|x_i|^2=\sum_{f=0}^{n-1}|X_f|^2$

Experimental Evaluation

Baseline methods:

- *FEDLwa_ANP*: All nodes participate in the federation process,
- *FEDLwa_RNP*: A randomly chosen portion (25%,50%,75% repsectively) of nodes participate in the federation process.

Network Configuration:

- Our testing took place in an ad hoc network consisting either of 20 or 50 nodes, respectively. The clustering algorithm was performed by setting d = 2.

Datasets:

- *Real dataset* (obtained by Kaggle) and consists of 52.585 temperatures (in Celsius degrees) of five cities in China for a time period of five years (2010-2015). We split our datasets for the purposes of training (chunks) and testing and specifically we keep the 90% of the dataset for training and the remaining 10% for testing.
- *Synthetic dataset*, which contains approximately 10.000 samples, which have undergone both erasion of the N U LL/ NaN values, standardization to meet the ML training standards and ARMA model processing in order for them to be converted into time series data.

Table 1: Statistics of the representative node at final chunk 27 and 20 for 20-node and 50-node network, respectively.

Algorithm	Loss		Mean Test Accuracy(%)
	20-node Network		
$FEDLwa\_ANP$	$\frac{epoch_{10_{th}}}{7.35}$	$\frac{epoch_{20_{th}}}{8.44}$	96.59
$L2-FEDLwa\_ANP$	11.76	9.12	96.73
$FEDLwa\_RNP_{25\%}$	29.67	32.83	81.47
$FEDLwa\_RNP_{50\%}$	42.31	35.57	74.99
$FEDLwa\_RNP_{75\%}$	50.708	46.64	72.38
$L2-FEDLwa\_RNP_{25\%}$	24.99	27.22	85.63
$L2-FEDLwa\_RNP_{50\%}$	48.95	38.61	74.72
$L2-FEDLwa\_RNP_{75\%}$	45.17	41.52	76.22
$\mathbf{FEDLwa\_SNP_{5coef}}$	9.88	7.28	**96.43**
$\mathbf{FEDLwa\_SNP_{10coef}}$	10.31	10.47	**97.26**
$L2-FEDLwa\_SNP_{5coef}$	19.03	16.24	96.45
$L2-FEDLwa\_SNP_{10coef}$	16.87	12.30	96.15
	50-node Network		
$FEDLwa\_ANP$	$\frac{epoch_{10_{th}}}{3.63}$	$\frac{epoch_{20_{th}}}{3.33}$	81.81
$FEDLwa\_RNP_{25\%}$	426,59	268,57	0.38
$FEDLwa\_RNP_{50\%}$	11.99	7.61	78.38
$FEDLwa\_RNP_{75\%}$	56.67	68.67	19.05
$\mathbf{FEDLwa\_SNP_{5coef}}$	87.06	72.27	**14.18**

Conclusions

In this work, we proposed a purely distributed federated deep learning methodology for small devices connected via wireless ad hoc networks, with the aim of overcoming both privacy-preserving problems and excessive communication costs. Towards these goals, we developed an ad hoc network clustering scheme, that creates relatively large clusters with the clusterhead appropriately place at the "center" of the cluster so as to avoid interference with nearby clusters, and also facilitate intercluster communication for carrying out the global averaging task. Local averaging takes place inside each cluster and a new node participation protocol into the federation was proposed, namely ***FEDLwa_SNP***. It is based on discovering similarity of the data among same cluster nodes by calculating small summaries that are transmitted to the CH which then decides which cluster members offer competitive ad- vantage to enter the federation. Results show that the proposed heuristic performs quite satisfactorily, compared to the *FEDLwa_ANP* baseline.

References

[1] E. T. M. Beltran, M. Q. Perez, P. M. S. Sanchez, S. L. Bernal, G. Bovet, M. G. Perez, G. M. Perez, and A. H. Celdran. 2023. ***Decentralized federated learning: Fundamentals, state of the art, frameworks, trends and challenges***. IEEE Communications Surveys & Tutorials 25, 4 (2023), 2983–3013.

[2] Evangelia Fragkou and Dimitrios Katsaros. 2024. ***Non-Static TinyML for Ad hoc Networked Devices.*** In TinyML for Edge Intelligence in IoT and LPWAN Networks, Bharat S Chaudhari, Sheetal N Ghorpade, and Marco Zennaroand Rytis Paškauskas (Eds.). Elsevier, 229–251. To appear.

[3] A. Oppenheim and R. Schafer. 2009. ***Discrete-Time Signal Processing.*** Prentice-Hall.

Figure 1: Random node performance, with id:18, after the global averaging procedure, taking place after the incoming data chunk 20, in 20-node network configuration.

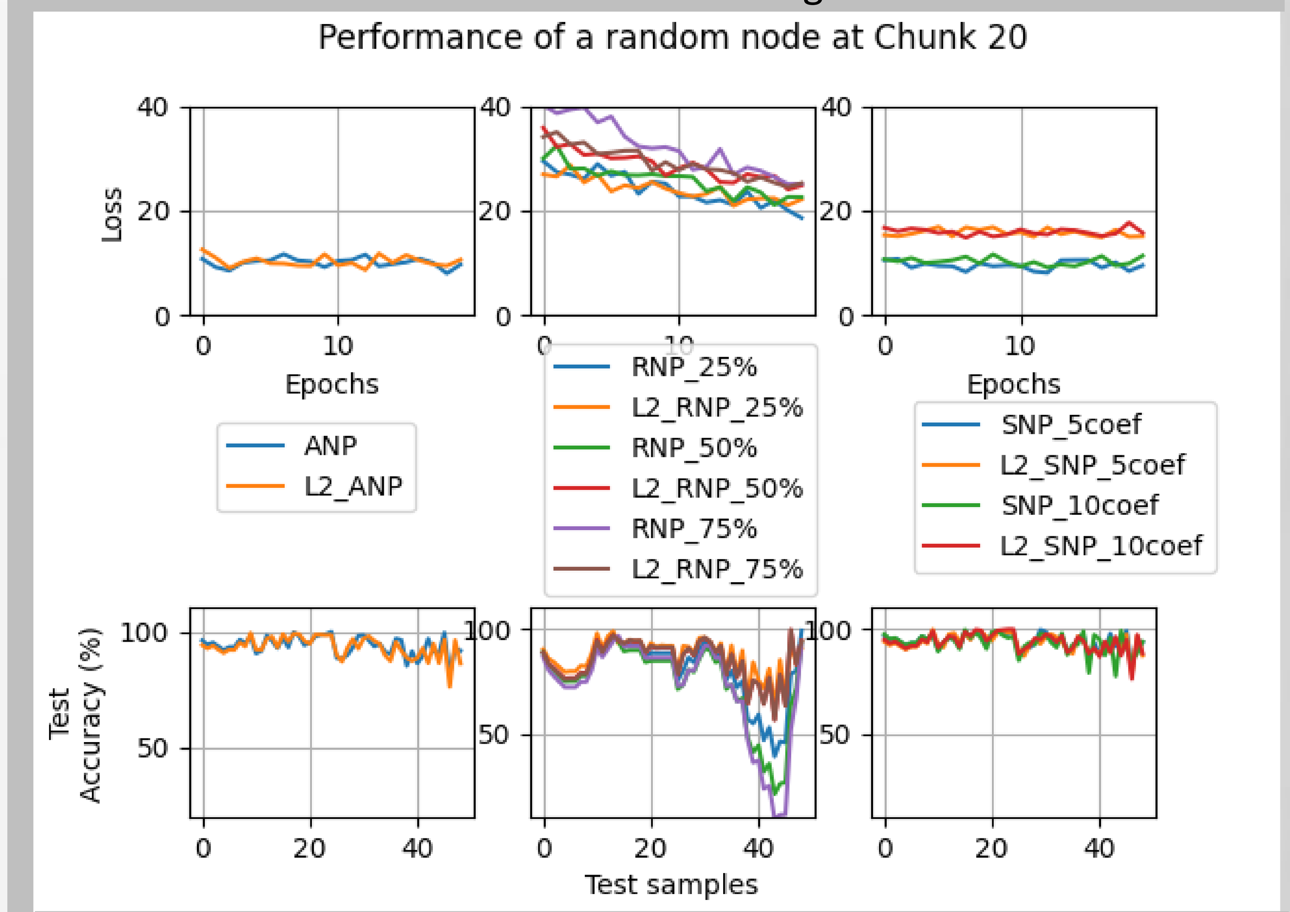
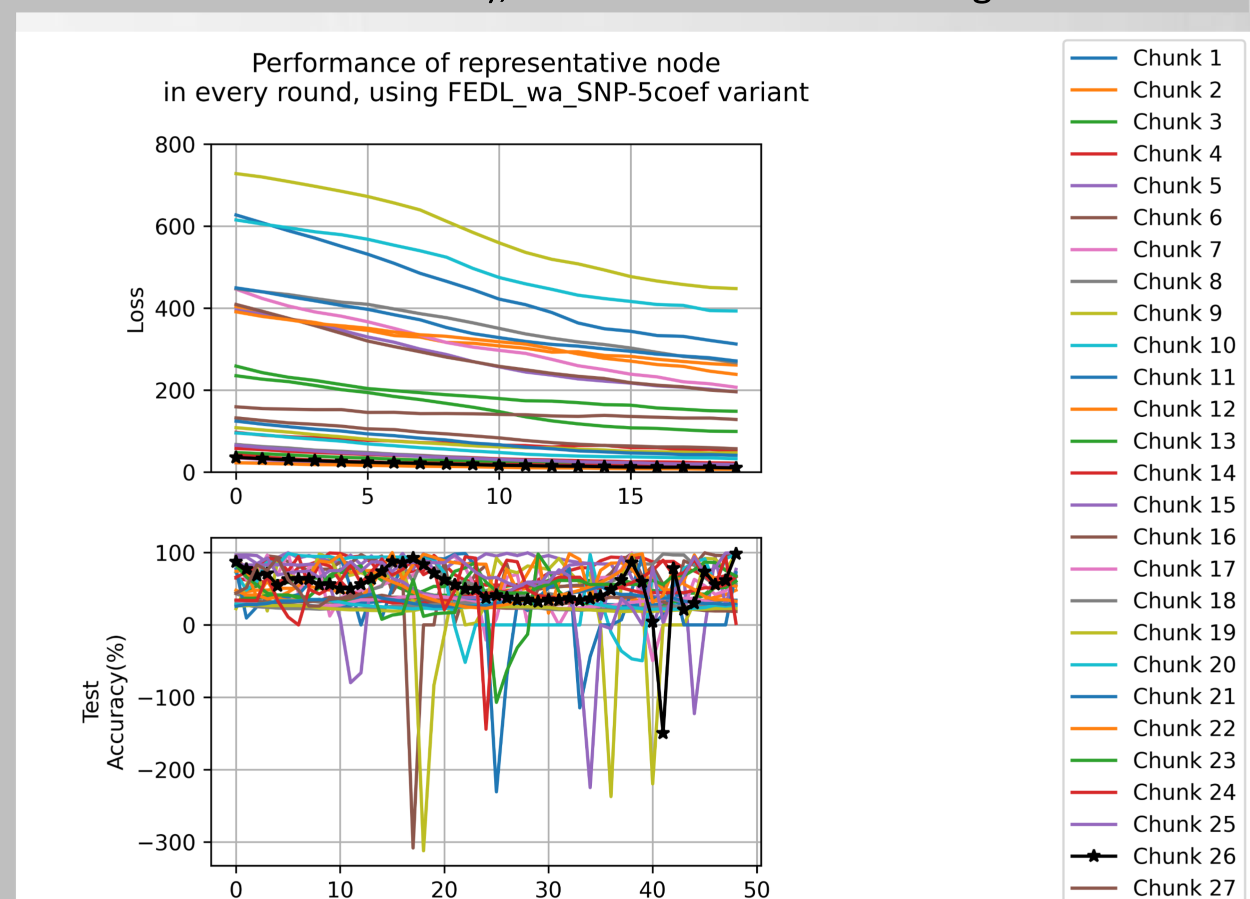


Figure 2: Progress of the representative node for the variants of ***FEDLwa_SN P***, in respect to the data chunks, given (emphasizing on the semi final one), in 20-node network configuration.



Acknowledgements

The research work is supported by the Hellenic Foundation for Research and Innovation (HFRI) under the 3rd Call for HFRI PhD Fellowships (Fellowship Number: *5631*).