University of Cagliari

DIEE - Department of Electrical and Electronic Engineering

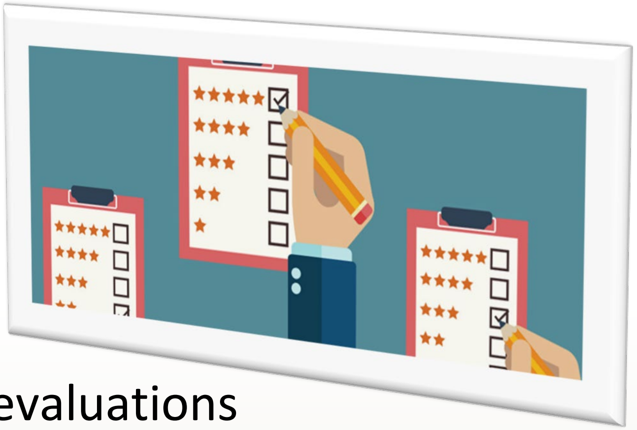
A CLUSTERED FEDERATED LEARNING APPROACH FOR ESTIMATING THE QUALITY OF EXPERIENCE OF WEB USERS

***Simone Porcu**, Alessandro Floris, Luigi Atzori*

Email: simone.porcu@unica.it

Context

- QoE estimation -> level of pleasure of the final user
  - Estimation using subjective assessments
  - Models to predict the perceived quality



- Issues?
  - Need for many subjective evaluations
  - People perceive the quality in different ways
  - Objective models are typically designed to estimate the Mean Opinion Score (MOS)
  - Privacy concerns: AI models need to centralise data

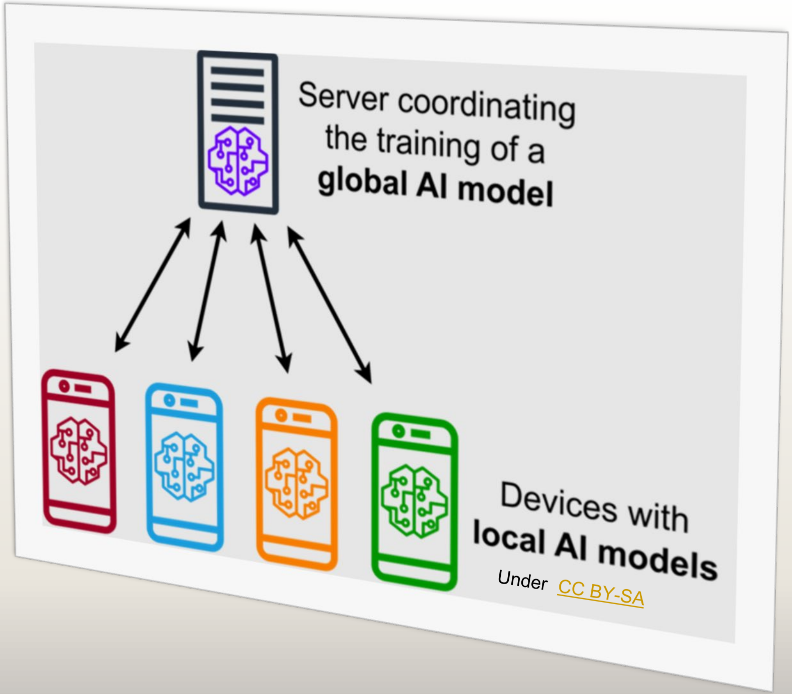
- Conclusion -> QoE models do not take into account the user's individual and personal differences and do not preserve their privacy

Federated Learning

- The Federated Learning (FL) paradigm provides an alternative solution to centralized-learning systems by adopting a collaborative learning approach where multiple entities (clients) collaborate in solving an ML problem under the coordination of a central server.

- FedAvg algorithm, collects the weights of the local models provided by the clients and computes the average of these weights to achieve a global model for all clients.

- QoE model personalisation?

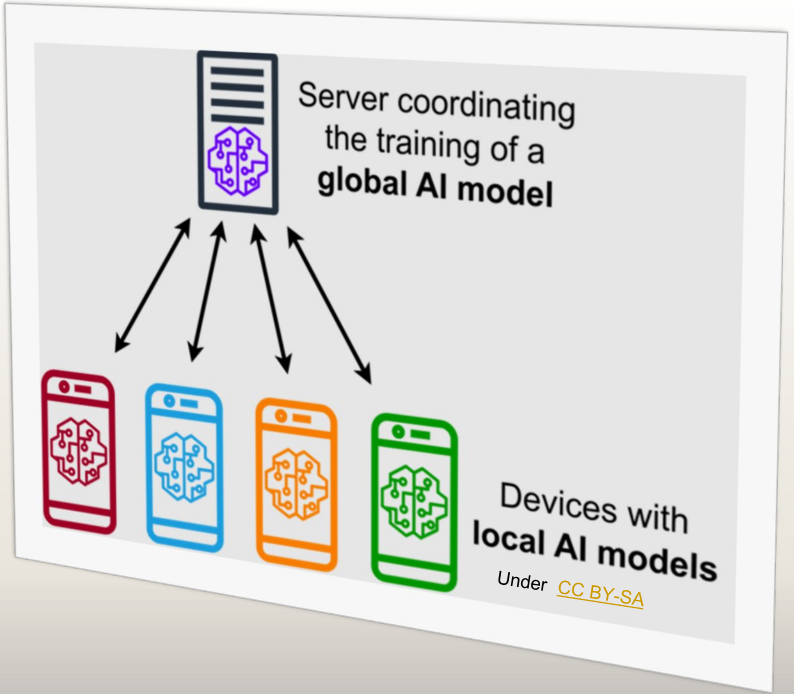


Federated Learning

- There are 100 users used to watch videos in 4K resolution and their perceived quality of that resolution is Excellent(5). If they watch a HQ (1920 × 1080) resolution video they will perceive a lower quality(3).
- There are other 25 users used to watch videos in HQ resolution, and their perceived quality of that resolution is Excellent (5).

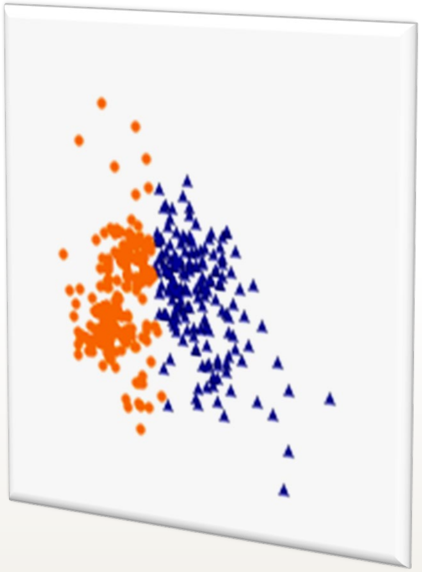
- The FL can not provide a good model for predicting the QoE

- The last 25 users will have a prediction of an average QoE value < 5



Clustering

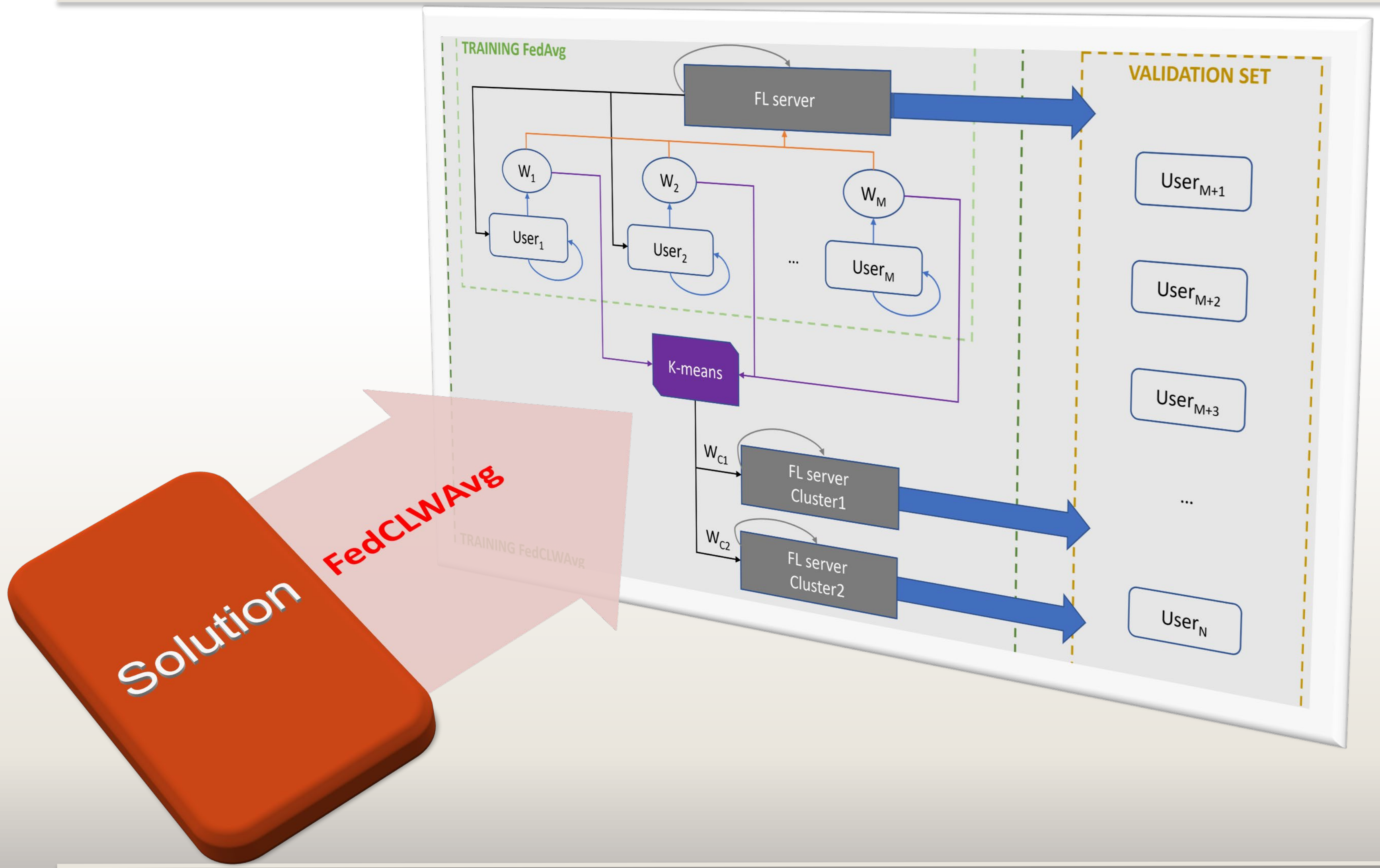
- To identify personal differences (in terms of perceived QoE for the same stimuli) between groups of users



- How can the clustering preserve the privacy?

- Which parameters should we use?
  - Personal evaluations are not privacy compliant

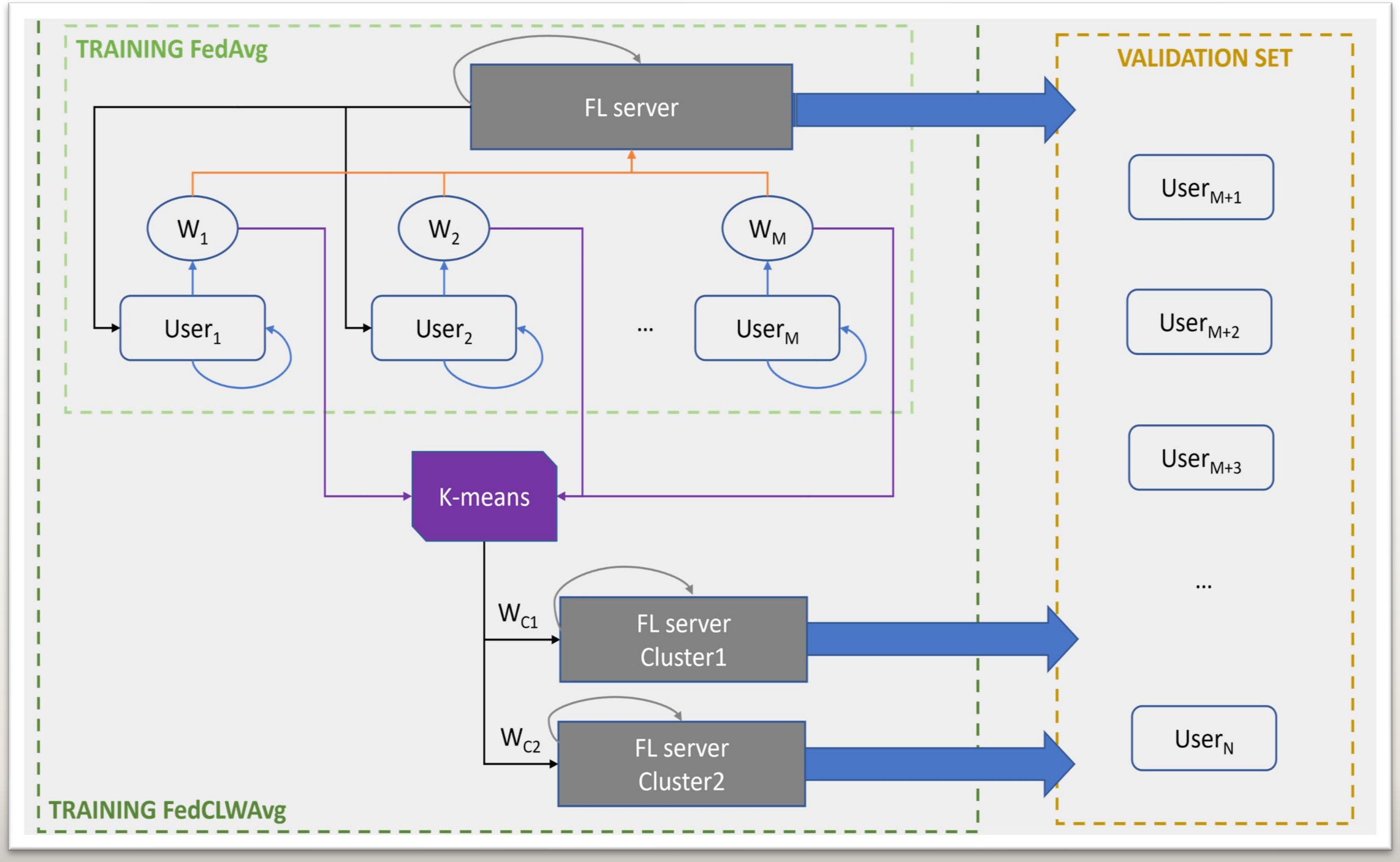
Proposed Solution



Proposed Solution

- FedAvg consist in:
  - Training NN on each client using its local data
  - Each client shares NNs weights to a FL server
  - The FL server averages the clients NN weights
  - The FL sends back to each client the averaged weight.
- FedCLWAvg is based on the hypothesis that personal differences (in terms of perceived QoE for the same stimuli) between groups of users are reflected in different weights of the trained local models
- FedCLWAvg consist in:
  - Run FedAvg training until each client reaches an accuracy of 0.70
  - If accuracy > 0.70, the FL server runs K-means on clients' weights
  - The K-means returns the number of identified clusters
  - New separate training sessions based on the FedAvg algorithm are initiated on each cluster.

Proposed Solution



Approach

- Demonstrate that is possible to create QoE models that respect the users' privacy.
- Demonstrate that is possible to create personalized QoE models that can overcome the perfomance of a MOS based neural network.

Use Case

- Prediction of the perceived QoE of *n* web users that provide feedbacks while browsing web-pages

Data

- Group of *135* web users provide feedbacks while browsing web-pages for a total of 3400 evaluations (from 1 to 5)*
- We selected the 9 features with the greatest Pearson Correlation Coefficient scores:
  - the time to load the Document Object Model (DOM)
  - the time to load the last visible image or other multimedia objects (Approximate Above-The-Fold, AATF)
  - the time to trigger the onload event (Page Load Time, PLT),
  - two ByteIndex (BI) metrics
  - two ObjectIndex (OI) metrics
  - two ImageIndex (II) metrics.

- Each client data has been augmented using ADASYN algorithm and normalized using the Z-score technique

* Dataset presented in D.N. da Hora, A.S. Asrese, V. Christophides, R. Teixeira, and D. Rossi, "Narrowing the gap between QoS metrics and Web QoE using Above-the-fold metrics," in Passive and Active Measurement (PAM), R. Beverly, G. Smaragdakis, and A. Feldmann, Eds. 2018, pp. 31– 43, Springer International Publishing.

Training

- We randomly removed the data related to ten users from the training dataset to be used as a validation set of the final trained NNs
- Every single user’s dataset was divided into training/validation subsets with a split rate of 70/30 for the training stage
- Both FedAvg and FedCLWAvg were trained for 2000 iterations in order to achieve comparable QoE estimation performance and convergence time.
- Concerning the FedCLWAvg approach, the computation of the K-means algorithm on the local weights returned 2 clusters (elbow method), which we call Cluster1 and Cluster2.

Results

Approach	Metric	Training ACR scores					Validation ACR scores				
		1	2	3	4	5	1	2	3	4	5
FedAvg	**Mean Accuracy**	0.83					0.71				
	Recall	0.99	0.99	0.54	0.79	0.83	0.83	0.70	0.63	0.40	0.99
	F1-Score	0.99	0.99	0.59	0.69	0.81	0.74	0.62	0.71	0.48	0.75
	Precision	0.99	0.99	0.59	0.69	0.81	0.74	0.61	0.71	0.48	0.75
FedCLWAvg Cluster 1	**Mean Accuracy**	0.89					0.80				
	Recall	0.83	0.99	0.90	0.84	0.90	0.99	0.63	0.57	0.99	0.80
	F1-Score	0.83	0.93	0.90	0.83	0.89	0.90	0.65	0.70	0.84	0.78
	Precision	0.92	0.92	0.90	0.83	0.89	0.90	0.65	0.70	0.84	0.78
FedCLWAvg Cluster 2	**Mean Accuracy**	0.86					0.82				
	Recall	0.99	0.99	0.71	0.60	0.99	0.99	0.76	0.87	0.66	0.82
	F1-Score	0.99	0.90	0.83	0.71	0.83	0.96	0.85	0.72	0.73	0.78
	Precision	0.99	0.90	0.83	0.71	0.83	0.96	0.85	0.72	0.73	0.78

Conclusion

- In general, the FL paradigm can avoid the Privacy issues that affect AI approaches

- The FedCLWAvg demonstrated that the clustering of the local weights could enhance the overall performance while preserving the users' privacy

- The FedCLWAvg can outperform the standard FedAvg approach

University of Cagliari

DIEE - Department of Electrical and Electronic Engineering

Thank you for the attention

Email: simone.porcu@unica.it