Machine Learning for Mobile Communication Systems

Problem for the Final Project

Statement

The dataset *D* stored in the file at this link reports the activity of one LTE cell in Madrid, El Rastro neighborhood. The observation period comprises the temporal window between the end of June and the beginning of August 2016 (06/29 - 08/09) that includes six weeks of information.
The dataset contains the following information:

- *overalltime:* in seconds, starting from the initial date of the measurements (29/06/2016)
- *overallusers:* number of transmitting users
- *overallrbdw:* total number of resource blocks in downlink
- *overallrbdwmean:* mean number of resource blocks in downlink
- *overallrbdwstd:* standard deviation of resource blocks in downlink
- *overallratedw:* rate (bit/s) in downlink
- *overallratedwmean:* mean rate (bit/s) in downlink
- *overallratedwstd:* standard deviation of rate in downlink
- *overallmsgdw:* number of messages in downlink
- *overallretxdw:* number of retransmitted packets in downlink
- *overallrbup:* total number of resource blocks in uplink
- *overallrbupmean:* mean number of resource blocks in uplink
- *overallrbupstd:* standard deviation of resource blocks in uplink
- *overallrateup:* rate (bit/s) in uplink
- *overallrateupmean:* mean rate (bit/s) in uplink
- *overallrateupstd:* standard deviation rate (bit/s) in downlink
- *overallmsgup:* number of messages in uplink
- *overallretxup:* number of retransmitted packets in uplink

Problem: let us split *D* in 6 portions ($D_1, D_2, D_3, D_4, D_5, D_6$), being one portion equal to one week (604800 seconds).
Given ($D_1, D_2, D_3, D_4, D_5, D_6$), it is requested:
1. for each D_i *(i=1, 2, 3, 4, 5, 6)*, build one model based on artificial neural networks able to automatically predict *overallusers*, *overallratedw*, *overallrateup*, at time instant *t+1*, *t+2*,…, *t+N*, with N=10. Select the most appropriate inputs to get the outputs requested with the lowest error. You can use one of the artificial neural network architectures introduced in the lessons (e.g., Multi-layer Perceptrons, Convolutional Neural Networks, Recurrent Neural Networks) or a combination of those.
2. create a federated global model using FedAvg algorithm, which is merging the local knowledge of the 6 datasets D_i and is performing the same task, i.e., prediction of *overallusers*, *overallratedw*, *overallrateup*, at time instant *t+1*, *t+2*,…, *t+N*, with N=10.
3. Evaluate the accuracy of the 4 local models and compare it with that achieved by the global model. For this, carefully identify the test sets to be used for the evaluation.

Optional: repeat 2 and 3 using FedProx algorithm. Comparison between FedProx and FedAvg may be included too.
Please, report all the model parameters used for the different simulations.

Instructions to deliver your project

What: a written report (paper style, two columns, pdf) including problem statement, proposed solution/method, achieved results and discussion. The generated code should be also delivered.
How: by email
When: 31 July 2024

Projects should be individual.

Project delivery, questions and comments: paolo.dini@cttc.es