Cluster-driven Graph Federated Learning over Multiple Domains

D. Caldarola, M. Mancini, F. Galasso, M. Ciccone, E. Rodolà, B. Caputo

Federated Learning

Federated Learning (FL) deals with learning a global model $M$ on the server-side in privacy-constrained scenarios, where data are stored on edge devices, i.e., the clients.


1. Federated Clustering of the local distributions, i.e. domains. Based on pseudo-labels predicted by the teacher–student domain classifiers.

   Test time: new domains as a soft combination of the previously discovered ones.

2. Cluster-specific models: domain specific parameters added as residuals in the model.

3. Graph-based interactions among domain-specific parameters.

Statistical heterogeneity in FL

In realistic scenarios, clients may hold different data distributions, leading to poor performances of the global model. Simply averaging the updates’ weights is not enough anymore.


<table>
<thead>
<tr>
<th>Dataset</th>
<th>Clients</th>
<th>Total samples</th>
<th>Samples per client</th>
<th>Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>CelebA</td>
<td>9,343</td>
<td>200,288</td>
<td>21.44</td>
<td>7,632</td>
</tr>
<tr>
<td>FEMNIST</td>
<td>3,550</td>
<td>805,263</td>
<td>226.83</td>
<td>88.94</td>
</tr>
</tbody>
</table>

Table 1: Datasets. Sources: LEAF benchmark.

Ablation studies

Experiments

We propose FedCG, a new algorithm for addressing statistical heterogeneity in FL, based on:

- an iterative clustering algorithm used to identify different data distributions and instantiate domain-specific parameters.
- Graph Convolutional Neural Networks for sharing knowledge across domains.
- the possibility to address new distributions at test time exploiting both the domain classifiers and the connections of the graph.


How to deal with non-i.i.d. and unbalanced clients’ data?

debora.caldarola@polito.it