Understanding Clinical Collaborations Through Federated Classifier Selection

Sebastian Caldas[1], Joo H. Yoon[2], Michael R. Pinsky[2], Gilles Clermont[2], Artur Dubrawski[1]

[1]Auton Lab, School of Computer Science, Carnegie Mellon University
[2]Department of Critical Care Medicine, School of Medicine, University of Pittsburgh

Let’s examine the advantages and obstacles of training an ML model on multiple clinical centers’ data.

- The advantages relate to the size of the data.
  - Larger sample sizes.
  - Increased availability of rare and new diseases.
  - Potential to enhance generalizability.
- The obstacles relate to accessing the data and deriving utility from it.
  - Limits on data sharing.
  - Population level differences may hurt model performance.
  - Overemphasis on predictive power.
    - We want to explain whether a decision is being made based mostly on external knowledge.

Let's examine the advantages and obstacles of training an ML model on multiple clinical centers' data.

- The advantages relate to the size of the data.
  - Larger sample sizes.
  - Increased availability of rare and new diseases.
  - Potential to enhance generalizability.
- The obstacles relate to accessing the data and deriving utility from it.
  - Limits on data sharing.
  - Population level differences may hurt model performance.
  - Overemphasis on predictive power.
    - We want to explain whether a decision is being made based mostly on external knowledge.

To address these obstacles, we propose FedeRated CLassifier Selection, or FRCLS.

Obstacle	FRCLS
Limits on data sharing.	Reuses classifiers trained in outside institutions, exchanging their associated parameters.
Population heterogeneity.	Adapts to the data distribution of each clinical center.
Overemphasis on predictive power.	Produce rules delineating regions of the feature space where the outside models outperform the local one.

To address these obstacles, we propose FedeRated CLassifier Selection, or FRCLS.

Obstacle	FRCLS
Limits on data sharing.	Reuses classifiers trained in outside institutions, exchanging their associated parameters.
Population heterogeneity.	Adapts to the data distribution of each clinical center.
Overemphasis on predictive power.	Produce rules delineating regions of the feature space where the outside models outperform the local one.

To address these obstacles, we propose FedeRated CLassifier Selection, or FRCLS.

Obstacle	**FRCLS**
Limits on data sharing.	Reuses classifiers trained in outside institutions, exchanging their associated parameters.
Population heterogeneity.	Adapts to the data distribution of each clinical center.
Overemphasis on predictive power.	Produce rules delineating regions of the feature space where the outside models outperform the local one.

To address these obstacles, we propose FedeRated CLassifier Selection, or FRCLS.

- We demonstrate the effectiveness of FRCLS between two different hospital systems on an early sepsis prediction task.

To address these obstacles, we propose FedeRated CLassifier Selection, or FRCLS.

- We demonstrate the effectiveness of FRCLS between two different hospital systems on an early sepsis prediction task.

Find more details at scaldas.xyz

Thank you!