UNIVERSITÀ DI PISA

An Approach to Federated Learning of Explainable Fuzzy Regression Models

José Luis Corcuera Bárcena, Pietro Ducange, Alessio Ercolani, Francesco Marcelloni, Alessandro Renda

University of Pisa, Dept. of Information Engineering, Pisa, Italy

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CROSSLAB Innovation for industry 4.0

Hexa-X

IEEE WCCI 2022 IEEE WORLD CONGRESS ON COMPUTATIONAL INTELLIGENCE PADUA, ITALY • 18 - 23 July, 2022

IEEE Computational Intelligence Society

Outline

- Introduction: motivation and objectives
- Fed-XAI: Federated Learning of eXplainable AI models
- From *traditional* TSK-FRBSs to *federated learning* of highly interpretable TSK-FRBSs
    - How to enforce interpretability in TSK-FRBS
    - How to address federated learning of TSK-FRBS
- Experimental setup and results

The pursuit of *trustworthiness*

Seven requirements toward Trustworthy AI

- Human agency and oversight
- Accountability
- Technical robustness and safety
- Societal and environmental wellbeing
- Diversity non-discrimination and fairness
- Privacy and data governance
- Transparency

Need to collect data to train accurate ML models clashes with need to preserve privacy of data owners.

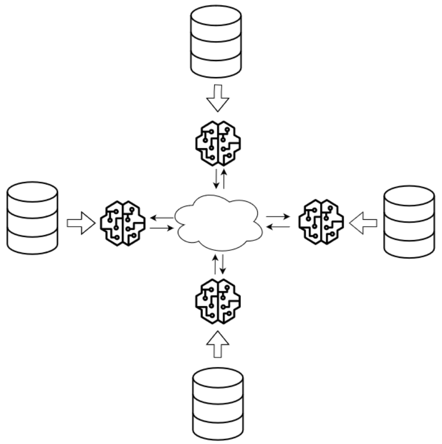
"AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned."

Fed-XAI

Federated Learning of eXplainable AI models

Fed-XAI: background

Federated Learning



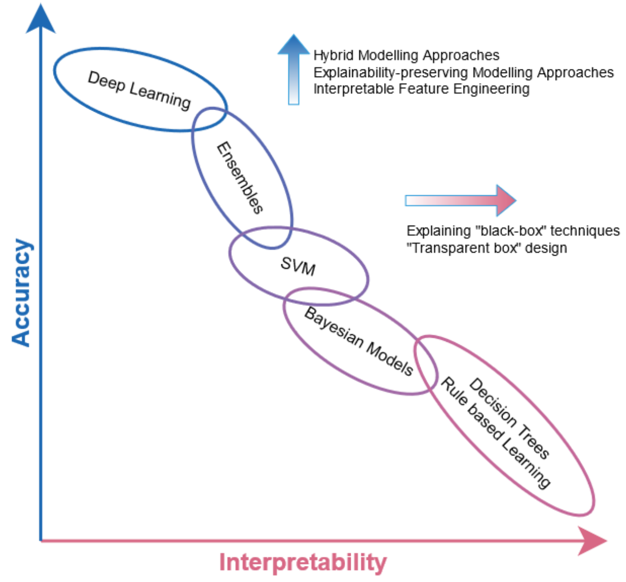
FedAvg (iterates over following steps):

1. server sends global model to clients
2. each client updates the model using local data and sends the model back to the server;
3. server takes the average of the locally computed updates, weighted according to the number of samples

- Suitable for models in which the learning stage is based on optimization of **differentiable global objective function** (e.g., NN)
- Ad-hoc strategies to be devised for other classes of models

eXplainable AI

- *Post-hoc* Explainability Techniques
- Design of Inherently Interpretable Models



Inspired to Arrieta et al. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." Information Fusion 58 (2020): 82-115.

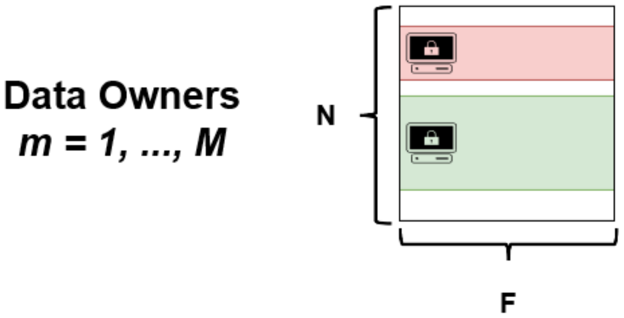
Objective: Federated Learning of TSK-FRBS

Objective

- Federated Learning of first-order **Takagi-Sugeno-Kang Fuzzy Rule-based systems** (TSK-FRBS)

Setting

- *Horizontally* partitioned data
- *M* data owners, *N* (overall) samples, *F* features



Contribution

- Design of an approach to **enforce interpretability** in first-order TSK-FRBSs
- Design of a novel approach for **aggregating** first-order TSK-FRBSs learned locally

The *traditional* TSK FRBS

Let

- $X = \{X_1, X_2, \dots, X_F\}$, be a set of input variable
- U_f, be the universe of discourse of variable X_f
- Y, be a continuous output variable
- $P_f = \left\{A_{f,1}, A_{f,2}, \dots, A_{f,T_f}\right\}$, be a fuzzy partition over U_f with T_f fuzzy sets

The generic k^{th} rule, R_k, of the rule base is in the form:

$$\textbf{IF } X_1 \textbf{ IS } A_{1,j_{k,1}} \dots \textbf{ AND } X_F \textbf{ IS } A_{F,j_{k,F}}$$

$$\textbf{THEN } y_k(\mathbf{x}) = \gamma_{k,0} + \sum_{i=1}^{F} \gamma_{k,i} \cdot x_i$$

Inference stage:

Given input pattern $\mathbf{x}$, compute strength of activation of each rule:

$$w_k(\mathbf{x}) = \prod_{f=1}^{F} \mu_{f,j_{k,f}}(x_f) \quad \text{for } k = 1,2,\dots,K$$

Estimation of antecedent parameters:

- Clustering in the input-output product space
- Fitting convex envelop of the projected membership values for each discovered cluster

Finally, generate the output as:

$$\hat{y}(\mathbf{x}) = \sum_{k=1}^{K} \left(\frac{w_k(\mathbf{x})}{\sum_{h=1}^{K} \left(w_h(\mathbf{x}) \right)} \right) \cdot y_k(\mathbf{x})$$

Estimation of consequent parameters:

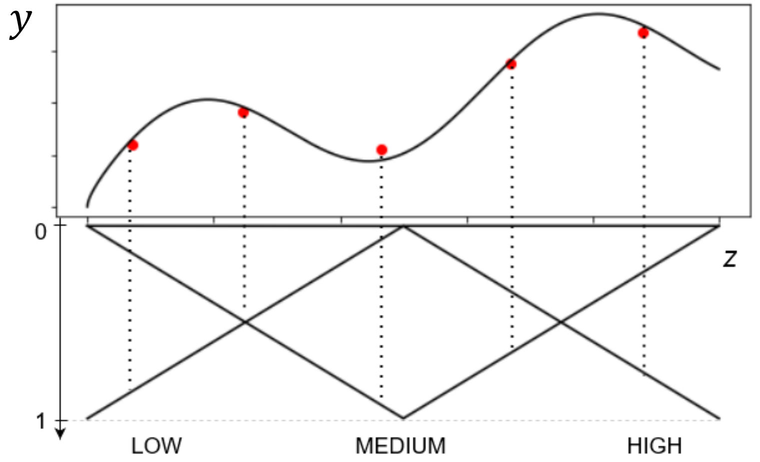
- Weighted Least Squared method

Enforcing Interpretability in TSK-FRBSs

Rules antecedents generation

1. **Strong triangular uniform** fuzzy partitioning on each normalized input attribute with $T_f = 3$ fuzzy sets
   → Coverage, completeness, distinguishability and complementarity (differently from "data-driven" partitions)
   → High semantic interpretability: «Low», «Medium», «High»

2. Numerosity reduction through **fuzzy clustering (FCM)** of training data in the input-output product space
   Generation of **antecedents based on centroids**
   → Summarization of training set and limited number of rules. Less complex models



Enforcing Interpretability in TSK-FRBSs

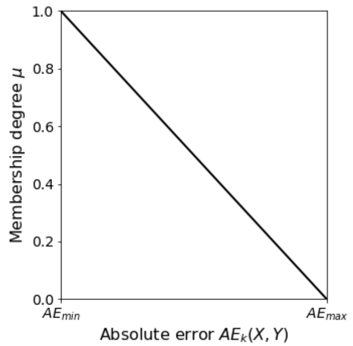
Inference process

- output determined by using rule with highest strength of activation (**maximum matching**)
  If more than one rule has same strength or no rule is activated, choose rule with highest *rule weight*

Rule weight RW_k of generic k^{th} rule, R_k

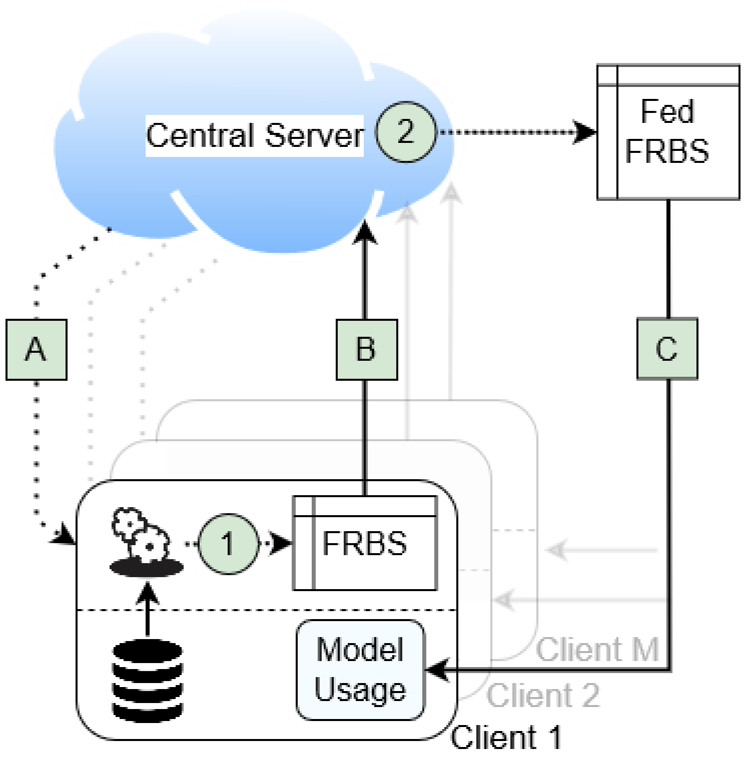
$$RW_k = 2 \cdot \frac{Supp_k \cdot Conf_k}{Supp_k + Conf_k}$$

$$Conf_k = \frac{\sum_{(x,y)\in TS} w_k(\mathbf{x}) \cdot \mu(AE_k(\mathbf{x},y))}{\sum_{(x,y)\in TS} w_k(\mathbf{x})} \qquad Supp_k = \frac{\sum_{(x,y)\in TS} w_k(\mathbf{x}) \cdot \mu(AE_k(\mathbf{x},y))}{N}$$



$$\mu(AE_k(\mathbf{x},y)) = 1 - \frac{AE_k(\mathbf{x},y) - AE_{min}}{AE_{max} - AE_{min}}$$

Our Federated TSK FRBS



A — **Configuration:** central server configures the learning process

1 — **Local learning of TSK-FRBSs**

B — **Transmission** of **local models** to the central server

2 — **Federated learning of the global TSK-FRBS:** aggregation of the models

C — **Transmission** of the **aggregated model** to the clients

Our Federated TSK FRBS – Aggregation Step

	Antecedent	*Consequent*	*Rule Weight*
	$ant_{1,1}$	$cons_{1,1}$	$rw_{1,1}$
	$\dots$	$\dots$	$\dots$
Client 1	$ant_{1,i}$	$cons_{1,i}$	$rw_{1,i}$
	$\dots$	$\dots$	$\dots$
	ant_{1,K_1}	$cons_{1,K_1}$	rw_{1,K_1}
$\dots$		$\dots$	
	$ant_{m,1}$	$cons_{m,1}$	$rw_{m,1}$
	$\dots$	$\dots$	$\dots$
Client m	$ant_{m,j}$	$cons_{m,j}$	$rw_{m,j}$
	$\dots$	$\dots$	$\dots$
	ant_{m,K_m}	$cons_{m,K_m}$	rw_{m,K_m}
$\dots$		$\dots$	
	$ant_{M,1}$	$cons_{M,1}$	$rw_{M,1}$
	$\dots$	$\dots$	$\dots$
Client M	$ant_{M,k}$	$cons_{M,k}$	$rw_{M,k}$
	$\dots$	$\dots$	$\dots$
	ant_{M,K_M}	$cons_{M,K_M}$	rw_{M,K_M}

Centralized server operation

1. Juxtaposition of rules collected from the M clients.
2. Identification of **conflicting rules:**
   (i.e., same antecedents, different consequents)
3. Replacement of conflicting rules with a new *single* rule:

- Antecedent: same of that of conflicting rules
- Consequent: coefficients computed as the weighted average of those from conflicting rules (weighted by RW)
- Rule weight (RW): average of rule weights of conflicting rules

The final rule base represents our Federated TSK model

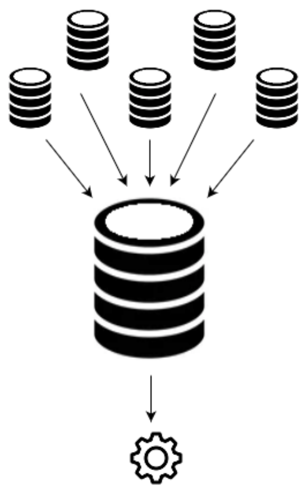
Experimental Setup

Dataset	Abbreviation	Dimensionality (F)	Samples (N)
Weather Izmir	WI	9	1461
Treasury	TR	15	1049
Mortgage	MO	15	1049
California	CA	8	20460

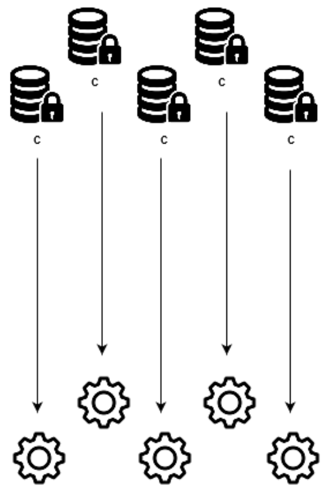
- Four regression datasets
- Params: $T_f = 3$, $C_{FCM} = 30$, $M = 5$
- Simulated distributed setting: randomly split each dataset (same number of instances) among **5 participants**

Three scenarios

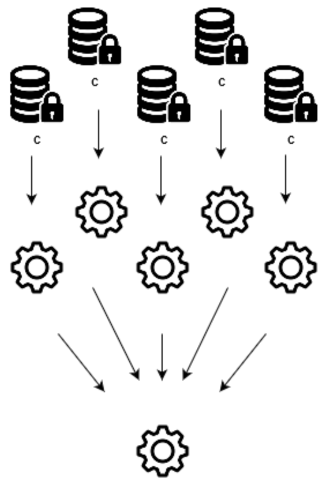
Centralized model: *no privacy*



Local model: *no collaboration*



Federated model: *privacy & collaboration*



Experimental Results

Setting

- **Mean Squared Error (MSE)** evaluated with **5-fold cross-validation**
- Each of the three scenarios evaluated on **same local splits**

Considerations

- *federated* always outperforms *local*, on average
- *federated* comparable to *centralized* for WI and CA
- *centralized* outperforms *federated* in case of high dimensionality $(F_{MO} = F_{TR} = 15)$ and data scarcity $(N_{MO} = N_{TR} = 1049)$
- performance comparable to those reported in the literature

Average MSE

Client ID	Local Train	Local Test	Federated Train	Federated Test	Centralized Train	Centralized Test
Weather Izmir						
1	1.33	*2.02*	1.44	*1.57*	1.40	*1.54*
2	1.09	*1.62*	1.25	*1.41*	1.22	*1.34*
3	0.96	*1.40*	1.25	*1.32*	1.22	*1.29*
4	1.07	*7.10*	1.23	*1.30*	1.20	*1.28*
5	1.19	*1.64*	1.41	*1.51*	1.38	*1.46*
Avg.	1.13	*2.76*	1.32	*1.42*	1.28	*1.38*
Treasury $(\times 10^{-3})$						
1	7.11	*377.40*	82.20	*112.72*	21.97	*46.13*
2	19.28	*192.70*	53.64	*79.41*	37.69	*51.35*
3	7.72	*337.25*	429.38	*174.18*	26.86	*41.97*
4	9.31	*110.47*	72.86	*378.61*	20.51	*41.69*
5	10.37	*133.83*	57.04	*40.85*	13.24	*20.37*
Avg.	10.76	*230.33*	139.02	*157.15*	24.06	*40.30*
Mortgage $(\times 10^{-3})$						
1	2.29	*78.08*	9.70	*15.96*	5.20	*7.55*
2	1.44	*15.08*	9.14	*7.35*	3.47	*5.22*
3	1.22	*38.18*	14.61	*9.52*	3.31	*5.22*
4	1.54	*53.84*	9.38	*35.90*	4.24	*8.83*
5	1.09	*43.36*	14.78	*5.14*	3.74	*4.98*
Avg.	1.52	*45.71*	11.52	*14.77*	3.99	*6.36*
California $(\times 10^{9})$						
1	4.73	*4.87*	4.75	*4.86*	4.77	*4.78*
2	4.62	*4.73*	4.57	*4.58*	4.60	*4.62*
3	4.71	*4.89*	4.71	*4.74*	4.72	*4.75*
4	4.77	*5.10*	5.23	*5.34*	5.18	*5.24*
5	4.70	*4.82*	4.63	*4.64*	4.65	*4.68*
Avg.	4.71	*4.88*	4.78	*4.83*	4.78	*4.81*

Experimental Results: additional considerations

Global interpretability as model complexity (average number of rules)

- data summarization strategy helps **limiting the overall number of rules**
- *local* and *centralized*: similar number of rules
- *federated*: generally more complex due to rule merging

Average number of rules

Dataset	Local	Centralized	Federated
Weather Izmir (WI)	13.96	13.40	27.80
Treasury (TR)	21.36	21.20	42.40
Mortgage (MO)	21.60	21.00	46.00
California (Ca)	8.80	8.60	10.20

Validation (centralized setting) of the proposed approach to learn **TSK-FRBSs with enforced interpretability**

- **TSK-SC:** our approach - single consequent (maximum matching)
- **TSK-AC:** our approach - averaging consequents (as in traditional TSK-FRBS)
- **pyFUME**: state of art approach (tuned at comparable complexity)

Average MSE

	TSK-SC		TSK-AC		PyFUME [5], [6]	
Dataset	Train	Test	Train	Test	Train	Test
WI	1.28	1.38	1.28	1.37	1.48	1.52
TR	24.06	40.30	24.42	39.18	32.07	62.93
MO	3.99	6.36	4.29	6.14	4.49	8.22
CA	4.78	4.81	4.82	4.85	4.62	4.64

Our **TSK-SC** achieves higher level of interpretability without compromising modelling capability

Fuchs et al., "pyFUME: a Python package for fuzzy model estimation," in 2020 IEEE Int'l Conf. on fuzzy systems
Fuchs et al. "Towards more specific estimation of membership functions for data-driven fuzzy inference systems," in 2018 IEEE Int'l Conf. on Fuzzy Systems

Conclusions

- Proposal of a novel **Fed-XAI solution**: Federated Learning of XAI models
  - TSK FRBSs slightly modified to achieve **high interpretability, without compromising performance**
  - **Aggregation** of first-order TSK-FRBSs learned locally in clients participating in the federation
  - Collaborative learning of ***federated*** TSK-FRBS model without disclosure of private raw data
- Preliminary experimental analysis
  - *Federated* approach **outperforms *local*** learning (no collaboration among clients)
  - *Federated* approach **outperformed by *centralized*** learning (unfeasible in privacy-sensitive applications)

- Main challenge to be addressed in future developments: how to tune the hyperparameters of our system?
  - number of clusters for data summarization
  - granularity of the fuzzy partitions

Thanks for your attention

Alessandro Renda

alessandro.renda@unipi.it

Department of Information Engineering

University of Pisa

Backup slides

Experimental Results

Results of the **Wilcoxon Signed-Rank test** on the MSE values obtained on the test sets

- *Federated* approach is selected as the control one and is compared with *local* and *centralized* ones
- Null hypothesis: the two approaches have the same level of performance
- Each distribution consists of **25** values of MSE measured on the test sets, derived from the iterations of the cross-validation over the involved clients

DS	R^+	R^-	p-value	Hypothesis ($\alpha = 0.05$)
Federated vs Local				
WI	314	11	0.0000	Rejected ($>$)
TR	230	95	0.0710	Not Rejected ($=$)
MO	309	16	0.0000	Rejected ($>$)
CA	237	88	0.0451	Rejected ($>$)
Federated vs Centralized				
WI	91	234	0.0551	Not Rejected ($=$)
TR	5	320	0.0000	Rejected ($<$)
MO	24	301	0.0000	Rejected ($<$)
CA	231	94	0.0667	Not Rejected ($=$)

Experimental Results

Execution time (seconds). Mean and standard deviation

- Communication times are not taken into account
- *Federated* approach runtime = slowest local training procedures + aggregation time
- Runtime of ***local*** and ***federated*** approaches are comparable
- Runtime of ***centralized*** case considerably higher

	Local	**Centralized**	**Fed. (Step 1)**	**Fed. (Step 2)**
WI	0.21 ± 0.02	1.22 ± 0.15	0.24 ± 0.02	$4.2e-3 \pm 6.1e-3$
TR	0.30 ± 0.03	1.51 ± 0.14	0.33 ± 0.02	$5.9e-3 \pm 5.9e-3$
MO	0.29 ± 0.02	1.45 ± 0.10	0.32 ± 0.02	$2.2e-3 \pm 0.2e-3$
CA	3.20 ± 0.70	24.19 ± 9.21	4.13 ± 0.58	$1.0e-3 \pm 0.1e-4$

Experimental Results

Model interpretability

- **antecedent** of a generic rule Rk identifies a specific region of the attribute space
- within this region, **predicted output** is evaluated as a **linear combination of input variables**
- **coefficient vector** describes the **effect of each attribute on the output value**

IF

longitude (x_1) is Low AND latitude (x_2) is Medium and housingMedianAge (x_3) is Medium AND totalRooms (x_4) is Low AND totalBedrooms (x_5) Low AND population (x_6) is Low AND households (x_7) is Low AND medianIncome (x_8) is Medium

THEN

$$\mathrm{medianHouseValue} = 0.83 - 1.08x_1 - 0.95x_2 + 0.08x_3 + 0.41x_4 + 2.18x_5 - 5.29x_6 + 0.27x_7 + 1.28x_8$$