

Federated Clustering Algorithm based on a Fuzzy Clustering Feature Vector

Asier Urio-Larrea

Universidad Pública de Navarra, Pamplona, Spain

asier.urio@unavarra.es

Introduction

The availability of a large amount of data produced in multiple devices and organizations provides great opportunities and challenges. A great amount of this data is privacy-sensitive and thus, sending it to a central server is not convenient. Furthermore, this data can have different characteristics in different producers making a unique globalized model not the best choice for its processing. Federated Learning is a concept developed to overcome these problems, through a local parameter learning step and a posterior parameter sharing step which does not involve sharing the data. In this work, a Federated Clustering algorithm is proposed. This algorithm is based on Fuzzy Cluster Feature vectors which can manage the uncertainty of points belonging to a certain cluster.

FCF Structure

The Fuzzy Cluster Feature Vector (FCF) structure's additive property enables the server to merge the compatible structures from different clients.

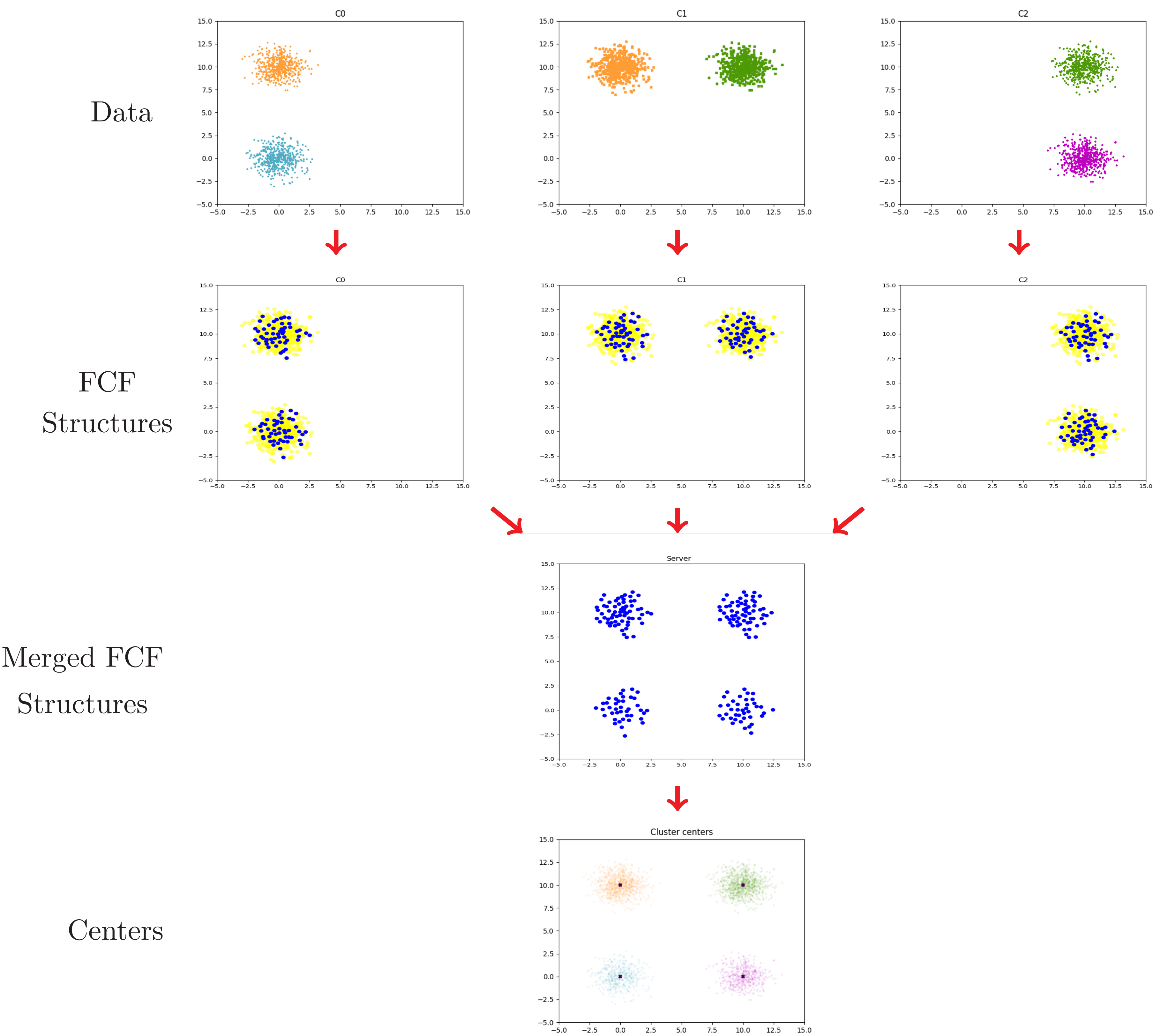
The fuzzy nature considers the membership (μ_j) of each point (x_i) to each structure (FCF_j). The components of the FCF structure are:

- $\overline{LS_j} = \sum_{x_i \in FCF_j} x_i \cdot \mu_j(x_i)$
- $SS_j = \sum_{x_i \in FCF_j} \|x_i - c_j\|^2 \cdot \mu_j(x_i)$
- $N_j = |FCF_j|$
- $M_j = \sum_{x_i \in FCF_j} \mu_j(x_i)$

FFCF Algorithm

- For Each** client c:
- For Each** point p:
- If** similarity(p,fcf) > threshold:
- merge(p,fcf)
- Else:**
- Create new FCF
- send FCFs to the server
- Server: merge received FCFs*
- For Each** FCF f:
- If** similarity(f, fcf) > threshold:
- merge(f, fcf)
11. Weighted FCM on the combined FCFs
12. Send back the combined FCFs and centers

Non-IID data example



Experimental results

Points per client	<i>FFCM</i>		<i>FFCF</i>	
	WSSE	OSSE	WSSE	OSSE
100 – 1000 – 100	0.63	17.18	0.62	17.05
100 – 1000 – 1000	0.63	17.21	0.61	17.11
1000 – 100 – 100	0.63	17.18	0.62	17.09
1000 – 1000 – 1000	0.62	17.14	0.61	17.10

Conclusions and Future Work

- The FFCF algorithm has a good performance in the studied case.
- It requires fewer communication rounds than the Federated Fuzzy C-Means.
- This demonstrates that the FCF structures are valid for a federated clustering application.

Further experimentation is needed:

- More datasets: both in IID and non-IID data distributions, synthetic and real.
- Test performance against a greater number of algorithms.
- Optimisation of the algorithm.

QR Code for Code

The source code is available on:
<https://github.com/asieriko/FFCF>



Aknowledgements

Supported by MCIN/AEI/10.13039/501100011033/FEDER, UE project PID2022-136627NB-I00 and by Contratos predoctorales Santander-UPNA 2021