International Cartographic Association
Association Cartographique Internationale

Spatio-Temporal Vertical Federated Learning to Overcome Data Sharing Limitations

Jose Antonio Lorencio Abril [a*], Anita Graser [a], Axel Weißenfeld [a], Anahid Wachsenegger [a]

[a] *Austrian Institute of Technology GmbH, [jose.lorencio-abril | anita.graser | axel.weissenfeld | anahid.wachsenegger]@ait.ac.at*

* Corresponding author

Keywords: Machine Learning, Federated Learning, GeoAI

Abstract:

Data are central to any Machine Learning (ML) application but often remain scattered in different parties' databases, hindering the development of effective and reliable models. The reluctance to share valuable data assets due to competitive concerns and strict privacy laws, such as the General Data Protection Regulation (GDPR) in Europe, adds complexity to data-sharing. This is further complicated when dealing with spatio-temporal data, which can potentially reveal individual identities through movement patterns when merged with other data sources, as shown in Rossi et al. (2015), creating a barrier to enhancing ML training processes through a broader data sharing.

Federated Learning (FL) has been proposed as a solution to address the challenge of data-sharing limitations by designing a secure way to collaboratively train an ML model without the need to share the raw data Yang et al. (2019). FL variants include Horizontal Federated Learning (HFL), which aims at obtaining ML models collaboratively from data partitioned in their sample space among different clients, and Vertical Federated Learning (VFL), in which the partition is in the feature space. FL could revolutionize location-based services that leverage GeoAI by enabling more effective information transfer without compromising user privacy.

We demonstrate the potential of FL to address a crowd prediction challenge. We propose a VFL framework in which we suppose that different clients hold different forecasting tasks and data relevant to each other. For example, one client may hold parking garage occupancy data and want to predict future occupancy levels, while another client may hold crowd density information in surrounding areas and want to predict crowdedness. Our proposed VFL setup works as follows:

1. **Setup.** Each active client holds two trainable ML modules: an *encoder*, and a *forecaster*. These modules can be different for each client, tailored to their needs. There can also be passive clients who don't have a forecasting need but collaborate by sharing information. Passive clients only have the *encoder* module.
2. **Data Alignment.** The parties need to align the data. Spatio-temporal data is usually aligned by time frames and locations, for example, through temporal and spatial binning (e.g. using discrete global grids or distance matrices).
3. **Training.** Each client:
   (a) Creates an encoded representation of their data using the *encoder*, and sends it to the rest of the clients.
   (b) Uses their own data and the encoded representations obtained from the other clients to make a prediction using the *forecaster* module.
   (c) Updates its *forecaster* according to a loss function suitable for its task, and sends the gradient obtained from the *forecaster* to the rest of the clients.
   (d) Aggregates its gradient with the gradients obtained from the rest of the clients, and updates its *encoder* module.

This way, clients can obtain a better model than possible by only using their own data while simultaneously keeping the training data private.

Previous work in this line of research is scarce, but in Li et al. (2024), the authors propose a collaborative framework to obtain a forecasting model without sharing clients' data by sharing latent representations of each client's data. However, their framework creates a single model that is useful for all clients, while our framework targets a more general scenario in which each client may have different forecasting needs. Other researchers have experimented with the application of Horizontal FL to spatio-temporal data with satisfactory results in anomalous trajectories detection (Koetsier et al. (2022)) and location recommendation (Rao et al. (2021)). Still, these studies assume a common feature space and forecasting objectives among clients.

Abstracts of the International Cartographic Association, 7, 95, 2024.
European Cartographic Conference – EuroCarto 2024, 9–11 September 2024, TU Wien, Vienna, Austria.
https://doi.org/10.5194/ica-abs-7-95-2024

Our crowd prediction problem has two active clients and one passive client. The first active client monitors the number of people in different areas of Scheveningen Beach, and the second active client monitors the occupancy levels of varying parking garages near this area. The passive client holds weather data that can be useful for the forecasting tasks of the active clients, while it does not have a particular forecasting task. Combining these datasets in the proposed VFL manner, we can increase the predictive performance of both active clients, who only need to share high-dimensional encoded representations of their data and gradients, making the process safe regarding data privacy.

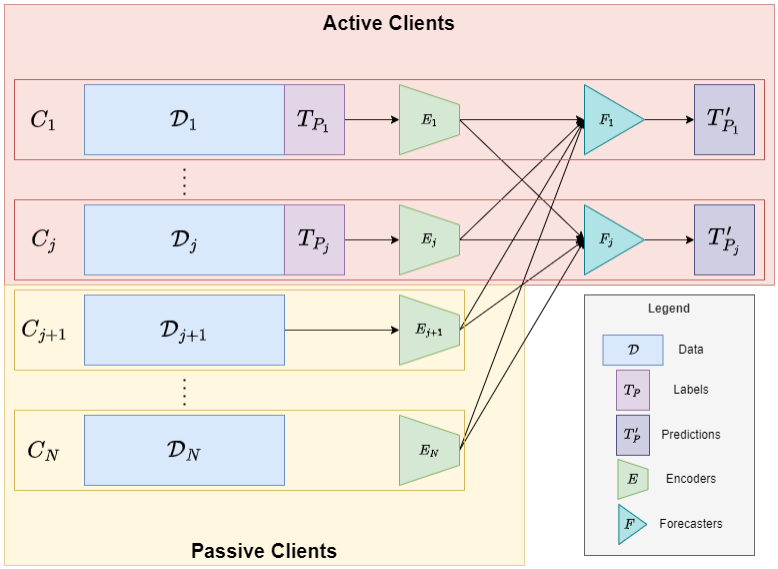

Figure 1. Proposed VFL framework.

We envision that this research can impact several areas of industry and research in the location-based services and mobility domain and beyond since it addresses the issue of incentives for joining a federation, which is a well-known problem that does not have a one-fits-all solution, Tu et al. (2022). Our approach increases the incentives of each party for collaboration by design since it allows each party to obtain a tailored private model that performs at a level that would not be attainable without the collaboration while, at the same time, keeping the training data private.

Acknowledgements

This work is funded by the EU's Horizon Europe Research and Innovation program under Grant No. 101093051 Emeralds.

References

Koetsier, C., Fiosina, J., Gremmel, J. N., Müller, J. P., Woisetschläger, D. M. and Sester, M., 2022. Detection of anomalous vehicle trajectories using federated learning. *ISPRS Open Journal of Photogrammetry and Remote Sensing* 4, pp. 100013. ADS Bibcode: 2022OJPRS...400013K.

Li, P., Guo, C., Xing, Y., Shi, Y., Feng, L. and Zhou, F., 2024. Core network traffic prediction based on vertical federated learning and split learning. *Scientific Reports* 14(1), pp. 4663. Publisher: Nature Publishing Group.

Rao, J., Gao, S., Li, M. and Huang, Q., 2021. A privacy-preserving framework for location recommendation using decentralized collaborative machine learning. *Transactions in GIS* 25(3), pp. 1153–1175. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/tgis.12769.

Rossi, L., Walker, J. and Musolesi, M., 2015. Spatio-temporal techniques for user identification by means of GPS mobility data. *EPJ Data Science* 4(1), pp. 11.

Tu, X., Zhu, K., Luong, N. C., Niyato, D., Zhang, Y. and Li, J., 2022. Incentive Mechanisms for Federated Learning: From Economic and Game Theoretic Perspective. *IEEE Transactions on Cognitive Communications and Networking* 8(3), pp. 1566–1593. Conference Name: IEEE Transactions on Cognitive Communications and Networking.

Yang, Q., Liu, Y., Chen, T. and Tong, Y., 2019. Federated Machine Learning: Concept and Applications. *ACM Transactions on Intelligent Systems and Technology* 10(2), pp. 12:1–12:19.

https://doi.org/10.5194/ica-abs-7-95-2024