

Personalized Federated Learning over non-IID Data for Indoor Localization

Peng Wu¹, Tales Imbiriba¹, Junha Park², Sunwoo Kim², Pau Closas¹

¹ Electrical and Computer Engineering Department, Northeastern University, Boston, MA
{wu.p,talesim,closas}@northeastern.edu

² Department of Electronic Engineering, Hanyang University, Seoul, Korea
{eric0725,remero}@hanyang.ac.kr

Abstract—Localization and tracking of objects using data-driven methods is a popular topic due to the complexity in characterizing the physics of wireless channel propagation models. In these modeling approaches, data needs to be gathered to accurately train models, at the same time that user's privacy is maintained. An appealing scheme to cooperatively achieve these goals is known as Federated Learning (FL). A challenge in FL schemes is the presence of non-independent and identically distributed (non-IID) data, caused by unevenly exploration of different areas. In this paper, we consider the use of recent FL schemes to train a set of personalized models that are then optimally fused through Bayesian rules, which makes it appropriate in the context of indoor localization.

Index Terms—Federated Learning, Bayesian inference, non-IID, data-driven, localization.

I. INTRODUCTION AND RELATED WORKS

There has been a steady increase in indoor real-time locating and tracking systems. With respect to the broader research domain of Positioning, Navigation and Timing, indoor positioning and tracking offers a number of difficult challenges. One approach to positioning and tracking in such environments is RF fingerprinting [1], [2]. The basic idea of fingerprinting is to build a database containing a collection of measured features at designated reference locations within the environment, and to then perform positioning by applying regression or classification techniques to match new measurements to one or more of those in the database. Acquiring the database is time consuming and costly, for which crowdsourcing approaches are relevant solutions. Traditional crowdsourcing techniques involved users sending their data directly to a centralized server, which in some situation may compromise users' privacy. Alternatively, Federated Learning (FL), has recently attracted great interest due to its privacy-protecting nature and the efficient use of resources by harnessing the processing power of edge devices [3]. In the context of localization, a number of studies have shown the potential of distributed learning [4]–[6].

FL is a promising solution that enables many clients to jointly train machine learning models while maintaining local data decentralization. Instead of exchanging data and conducting centralized training, each party sends its model to

the server, which updates a joint model and sends the global model back to the parties. Since their original data will not be exposed, FL is an effective way to address privacy issues [7].

A key and common data challenge in such distributed approaches is that the data from different clients is often non-independent and identically distributed (non-IID). For instance, this is the case when different FL clients collect data from distinct locations. According to previous studies [8]–[10], non-IID data settings (such as label distribution skewness, feature distribution unbalance, or training data amounts) reduce the effectiveness of learning [11].

Recent efforts to address FL under non-IID data have been proposed in [10], [12], [13] but considering a single global objective. However, it is difficult to converge to a global model that provides a good personalized performance on every client when the data is non-IID across different clients. Another branch of studies [11], [14]–[18] have addressed the need of personalized local models for each client. Most of the methods [15], [16], such as transfer learning or meta-learning, adapt or adjust the global model for individual clients, while some methods [17], [18] utilize pair-wise collaboration among clients to train personalized models.

In this paper we envisage the use of personalized FL models for localization purposes. In particular, we consider crowdsourced received signal strength (RSS) of WiFi data to build fingerprinting (data-driven) models, an approach that does not compromise training user's privacy. This problem connects with several scenarios where clients might be moving in large, and possibly intricate, areas such as market halls, hospitals, university campus, or even entire cities. In such scenarios the non-IID nature of the collected data is evident, which is illustrated in Fig. 1 with different clients focusing on different areas. The inherent complexity of local buildings and signal propagation makes personalized models appealing solutions that can obtain region-specific models and at the same time exploit information obtained from other areas. For this reason, we leverage Personalized FL strategies formulating the localization problem as a classification problem. We specially focus on the recent algorithm FedAMP [17] since it shows superior classification performance in comparison to related algorithms. Although FedAMP obtained good performance on different classification problems the direct application to

This work has been partially supported by the NSF under Awards CNS-1815349 and ECCS-1845833.



Fig. 1. Personalized FL involves clients (or agents) exploring an area in a possibly uncoordinated manner with the goal of gathering features of interest to train local models. As a consequence, different clients might observe non-IID data due to exploring different areas more often. In the figure, different markers represent data collection from different agents.

localization problems requires some adjustments. This is so, due to the data association problem between data acquired in an unknown location and all possible personalized models. To address this issue, we propose a Bayesian fusion strategy that combines the result of multiple personalized models into a unique global model during the test phase.

II. FEDERATED LEARNING SCHEMES

The idea behind the FL is to approximate a global model as an integration of local models. Mathematically it can be formulated as the minimization of a loss function:

$$\min_{\mathbf{W}} \mathcal{L}(\mathbf{W}) \text{ where } \mathcal{L}(\mathbf{W}) = \sum_{i=1}^M \mathcal{F}_i(\mathbf{w}_i) \quad (1)$$

where $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_M] \in \mathbb{R}^{d \times M}$ contains the model parameters, $\mathcal{L}(\mathbf{W})$ is the global cost functional, while $\mathcal{F}_i : \mathbb{R}^d \rightarrow \mathbb{R}$, $\mathbf{w}_i \mapsto \mathcal{F}_i(\mathbf{w}_i)$ are local cost functions trained using disjoint datasets $\mathcal{D}_1, \dots, \mathcal{D}_M$.

Among the different strategies to solve (1) we highlight the conventional FL approach: FedAvg, which considers a single global objective, as well as the latest personalized approach FedAMP, which considers several personalized local models.

A. FedAvg

Federated averaging (FedAvg) is a *de facto* approach for FL in which local models are trained locally. The clients then upload their models to a cloud server, which is in charge of fusing the model parameters (e.g. weights in a neural network) to compute a global model. In order to account for each local model quality, (1) can be modified as:

$$\mathcal{L}(\mathbf{W}) = \sum_{i=1}^M \frac{N_i}{N} \mathcal{L}_i(\mathbf{w}_i) \quad (2)$$

where $\mathcal{L}_i(\mathbf{w}_i) = \frac{1}{N_i} \sum_{n \in \mathcal{P}_i} f_n(\mathbf{w}_i)$, $f_n(\mathbf{w}_i)$ is the loss of the prediction on sample n made with model parameters \mathbf{w}_i . \mathcal{P}_i is the data partition for client i , N is the total number of data points and N_i is the number samples used by the i -th client. FedAvg is an efficient FL scheme, but may lead to poor

accuracy results in certain situations such as non-IID cases, as shown in previous studies [8], [9].

B. FedAMP

The federated attentive message passing (FedAMP) algorithm was devised as a privacy-preserving personalized cross-silo federated learning strategy for non-IID data [17]. It aims at solving an optimization problem of the form:

$$\min_{\mathbf{W}} \mathcal{L}(\mathbf{W}) + \lambda \mathcal{R}(\mathbf{W}) \quad (3)$$

where $\lambda \in \mathbb{R}_+$ and $\mathcal{R}(\mathbf{W}) = \sum_{i \neq j} h(\|\mathbf{w}_i - \mathbf{w}_j\|^2)$ is a regularization function composed of continuously differentiable attention-inducing functions $h : [0, \infty] \rightarrow \mathbb{R}$ aimed at constraining the M models to have *similar* parameters. The problem in (3) was reformulated by exploiting incremental gradient proximal methods to sequentially estimate the personalized model parameters \mathbf{W} , by solving

$$\mathbf{W}^k = \arg \min_{\mathbf{W}} \mathcal{L}(\mathbf{W}) + \tilde{\lambda} \|\mathbf{W} - \mathbf{U}^k\|_F^2 \quad (4)$$

at every iteration $k = 1, \dots, K$, where $\mathbf{U}^k = [\mathbf{u}_1^k, \dots, \mathbf{u}_M^k]$ is used as the prox-center [17], and can be updated as

$$\mathbf{U}^k = \mathbf{W}^{k-1} - \alpha_k \nabla \mathcal{R}(\mathbf{W}^{k-1}). \quad (5)$$

In (5) $\alpha_k \in \mathbb{R}_+$ is the step size of the gradient descent method and $\tilde{\lambda} = \lambda / 2\alpha_k$, where $\lambda \in \mathbb{R}_+$ is a parameter controlling the cross-silo regularization. It has been shown [17] that $\mathbf{u}_i^k \in \mathbb{R}^d$ computed in (5) can be written as a convex combination of the model parameters as:

$$\mathbf{u}_i^k = \xi_{i,1} \mathbf{w}_1^{k-1} + \dots + \xi_{i,M} \mathbf{w}_M^{k-1} \quad (6)$$

where $\xi_{i,i'} \in \mathbb{R}_+$, with $\sum_{i'} \xi_{i,i'} = 1$, can be interpreted as similarity coefficients due to the properties of the attention function h [17].

Important features of the methodology proposed by Huang et al. [17], briefly discussed above, include privacy preserving, since data used to learn local models are never shared with other nodes, and collaborative learning using an attentive message passing mechanism which ultimately induces local models with high similarity coefficient to collaborate. These features are extremely important for distributed indoor localization problems since data can be scarce and privacy is crucial, specially in scenarios where local models from different areas (or buildings) might be exchanging information. This is even more so, if considering large scale scenarios where models might be learned over multiple neighbourhoods or even over entire map.

Although this strategy proved relevant for different classification problems [17] its direct application to indoor positioning is not so direct, presenting challenges regarding data distribution and model prediction. Regarding data collection and distribution we will assume that data used to train a particular model can come from any location but will be more concentrated at the region associated with a given local model. Regarding the prediction step, since data arriving comes from an unknown location we propose to fuse multiple local models

to provide a final prediction. This will be addressed in more detail in the next section.

III. FUSION OF FEDERATED LEARNING MODELS

The FedAMP approach discussed earlier produces a set of M models that are well fitted to the data they were trained with. That is, if data was primarily from certain areas in the map, the predictive accuracy of those models is expected to be larger in those areas. Conversely, worse in those areas where data quantity is smaller than other personalized models. As a consequence, a user utilizing the models has no a priori knowledge which is the correct model to use for localization purposes (that is precisely the objective of the models). Therefore, we propose a fusion scheme that blends the M available models to produce a localization estimate. Here we are considering that the localization task involves determining which area (e.g. room, hall, etc.) the user is in, given the measured RSS data.

In the FedAMP scheme, the $i \in \{1, \dots, M\}$ classifier (i.e. each of the personalized models) provides a categorical posterior distribution for c , the class, which can take any of the $j \in \{1, \dots, L\}$ labels (i.e. each room or area) based on the current data \mathbf{y} and the corresponding model \mathcal{M}_i , that is

$$p(c|\mathcal{M}_i, \mathbf{y}) = \prod_{j=1}^L p_{j|i}^{[c=j]}, \quad (7)$$

where $p_{j|i}$ denotes the probability for the j -th label given the i -th classifier, and $[c=j]$ is an indicator function that returns 1 if $c=j$ and 0 otherwise. The a priori class probability $p(c)$ is categorical and is defined by the probabilities $p_{j|0}$, which in the equiprobable case result in $p_{j|0} = 1/L$, $\forall j$. The optimal fusion rule is provided by the joint posterior distribution, which can be shown [19] to be proportional to

$$\begin{aligned} p(c|\mathcal{M}_{1:M}, \mathbf{y}) &\propto \frac{\prod_{i=1}^M p(c|\mathcal{M}_i, \mathbf{y})}{p(c)} = \frac{\prod_{i=1}^M \prod_{j=1}^L (p_{j|i})^{[c=j]}}{p(c)} \\ &= \prod_{j=1}^L \left(\frac{p_{j|1} \cdot p_{j|2} \cdots p_{j|M}}{p_{j|0}^{M-1}} \right)^{[c=j]}, \end{aligned} \quad (8)$$

where we have used that the different models are conditionally independent given c . The resulting joint distribution is categorical, from which the maximum a posteriori (MAP) can be readily obtained to predict the class c .

IV. POSITIONING EXPERIMENTS

For our experiments we use the public available RSS data from UJIIndoorLoc database [20] is used to evaluate our proposed federated learning-based localization system. The UJI-IndoorLoc database includes 19937 RSS samples distributed over 390 meters in length to 270 meters in width area over four floors for training and testing. In our experiments we focus on floor 1 located at building 1 to reduce the computational cost. Figure 2 shows the latitude and longitude of the rooms in the selected floor. Because some rooms had very few samples, we

grouped neighbouring rooms to form larger areas depicted in Figure 2 by color and labeled from 0 to 9. Our experiments will focus in classifying RSS data coming from these 10 areas, but first we will present our non-IID data splitting strategy.

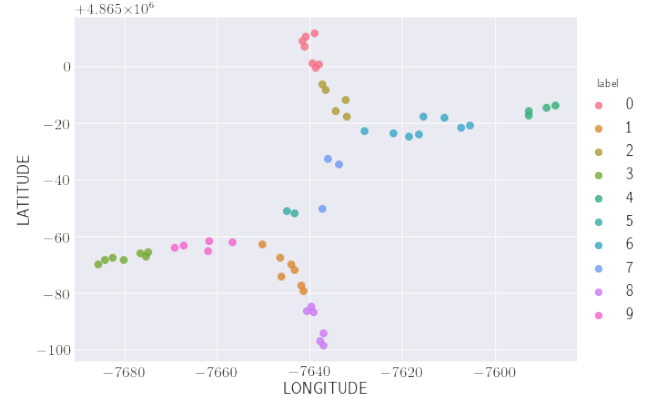


Fig. 2. Latitude/Longitude map of the data location in the floor under test. Different color represent clusters of nearby data, treated as a single label in this study.

A. Settings of non-IID Datasets

Among different non-IID possibilities, we focus here on having an unbalanced distribution of the labels. For example, some clients collect more specialized points in several specific areas after navigating the scene more often. To simulate label distribution skewness across clients, we use the Dirichlet distribution. This method has been used in many recent FL studies [11]. Specifically, for a given client i , we define the probability of sampling data from the $j \in \{1, \dots, L\}$ label as the vector $(p_{i,1}, \dots, p_{i,L}) \sim \text{Dir}(\beta)$, where $\text{Dir}(\cdot)$ denotes the Dirichlet distribution and $\beta = (\beta_{i,1}, \dots, \beta_{i,L})$ is the concentration vector parameter ($\beta_{i,j} > 0, \forall j$). The advantage of this approach is that the imbalance level can be flexibly changed by adjusting the concentration parameter $\beta_{i,j}$. If it is set to a smaller value, the partition is more unbalanced.

However, unlike other works, we are not only interested in generating some unbalanced data proportion, but also want that some clients have a similar label distribution, similarly to [17]. Fig. 3 shows an example non-IID distribution with a total of 10 labels and 10 clients gathering data in the area. A total of 3 groups are considered, each group containing 3 dominating labels. In this setup, we assign the dominating classes in first group by sampling data for clients 0 – 2 using large β values in those labels $\beta = (80, 80, 80, 20, \dots, 20)$. Similar strategy is used when assigning the dominating classes in other groups of clients (3 – 5 and 6 – 9 in the example). Using the aforementioned Dirichlet sampling approach provides more randomness to the study, as opposite to setting a fix number of data points per client.

B. Results

In this section we compare five different strategies that differ in their training and testing settings. Specifically, we

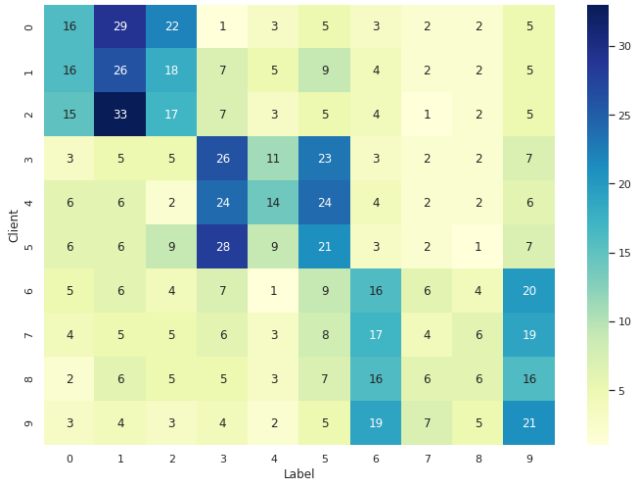


Fig. 3. An example of distribution label imbalance partition on the dataset. The value in each rectangle represents the number of data samples of a given label belonging to a given client. In this example, 3 groups of clients can be observed, each having similarly distributed data but non-IID across groups.

consider Global Model (GM) and Local Models (LMs) where the first is trained using data from all areas whereas for LMs each individual model is trained using only non-IID data associated to its specific area(s). Those models are neural networks (NN) with the same MLP structure, having two hidden layers with [256, 16] neurons and ReLU activations. Furthermore, as discussed in Section III we perform fusion of the LMs class predictions to generate a unique prediction. We call this LM with Fusion (LM-F). As a comparison we also present results when LMs are directly used for classification of RSS values. In this case the accuracy is computed averaging the accuracy obtained for each local model. The same applies for the FL models FedAvg, FedAMP and the fused version of the latter: FedAMP-F. Notice that FedAvg does not have fused version because it is already based on building a unique global model. For all experiments we performed 50 Monte Carlo runs (randomly initializing the NNs and performing data shuffling) to make the results more stable and reliable.

Figure 4, shows how the probability distribution changes before (FedAMP) and after fusion (FedAMP-F) is performed. Specifically, when class $c = 1$ is to be classified, the top panel shows the probability distribution of $p_{1|i}$, $i = 1, \dots, M$. We can observe a large performance variability depending on the used model. The bottom panel shows the distribution of $p(c = 1 | \mathcal{M}_{1:M}, \mathbf{y})$ after fusion, which shows clear improvements in terms of classifying correctly the label.

We present performance results in Figures 5 and 6 where GMs are used merely for benchmarks purposes. Fig. 5 shows the accuracy of all methods under the non-IID data setting and different number of classes L while keeping the number of clients fixed $M = 6$. Analyzing the figure we observe that, for all models, the performance decreases as the number of classes increases. This is expected since increasing L reduces the number of training samples available for each region, N_i . When comparing the different non-global strategies we

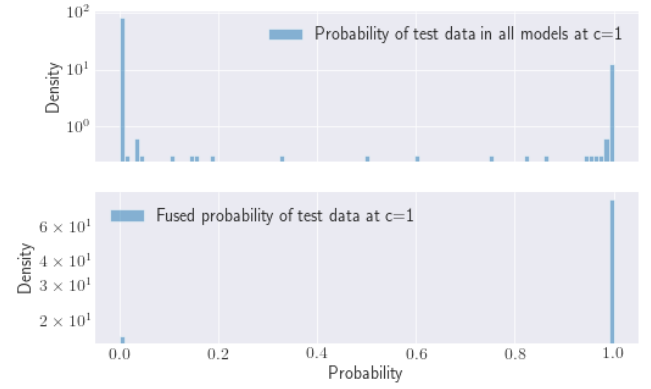


Fig. 4. Distribution of classification probabilities (top) before and (bottom) after fusion of the different local models.

highlight the poor accuracy of LM models. This is so, due to the data association problem and the lack of collaboration among clients. However, large improvements are achieved when introducing client collaboration (i.e., FL methods) or/and fusion. For instance, LM-F produces an average accuracy improvement in comparison with LM of approximately 18.7% while FedAvg and FedAMP leads to average improvements of 23.4% and 21.7%, respectively. The best local model results were obtained when considering FL strategies in conjunction with the Fusion strategy. This can be seen for all number of classes where FedAMP-F presents the best non-global result. In fact, for $L = 10$ and $L = 15$ FedAMP-F performs even better than the GM. This is due, however, to the fact that the same NN structure is used for GM and LMs giving local models more flexibility. In Fig. 6 we also examine the

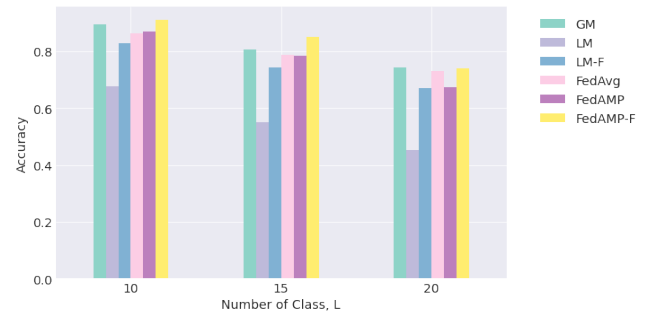


Fig. 5. Accuracy of models as a function of L , the number of classes.

performance variation across different number of clients, M , while keeping the number of classes fixed, $L = 10$. As the number of clients increase, the accuracy of LMs and LM-Fs decrease. This behavior is also expected since increasing the number of clients also reduces the amount of data available for training each client. While FedAvg achieves relatively stable results, the accuracy of FedAMP is decreasing as M increases. However, fusing the local FedAMP models (FedAMP-F) leads to the best results.

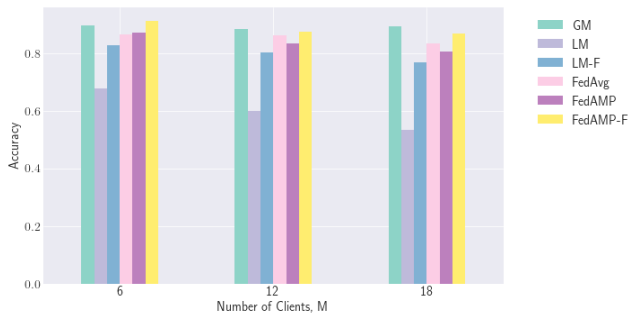


Fig. 6. Accuracy of models as a function of M , the number of clients.

C. Results for different hyper-parameters settings

In this section we examine the performance of the FedAMP and FedAMP-F as function of their hyperparameters: $\tilde{\lambda}$ and σ . Fig. 7 shows relative performance curves where “Rate” is the ratio between the accuracy of FedAMP-F, GM and FedAvg as described in the corresponding legends. Observing the top panel we can see that FedAMP-F seems quite insensitive to variations of σ . The bottom panel, however, shows great impact on the FedAMP-F performance when varying $\tilde{\lambda}$. The best results were obtained in the range $[0.01, 1]$. The results in Figures 5 and 6 were obtained with $\sigma = 20$, $\tilde{\lambda} = 1$.

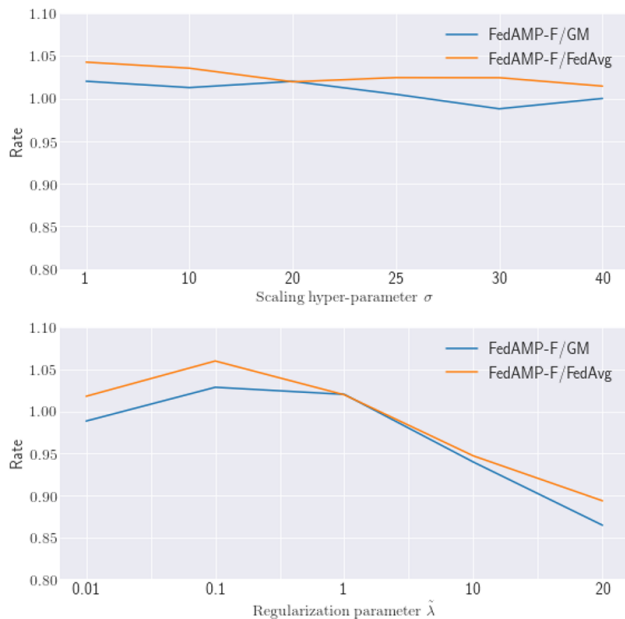


Fig. 7. Sensitivity analysis of FedAMP-F to (top) σ and (bottom) $\tilde{\lambda}$.

CONCLUSIONS

In this paper, we tackle the challenging problem of non-IID data in indoor positioning by proposing a Bayesian fusion rule to merge personalized Federated Learning models. This solution enables the applicability of such training schemes in situations where we do not have prior knowledge regarding which model to use. We show how the fused model can

remarkably improve the performance with respect to personalized models.

REFERENCES

- [1] D. Dardari, P. Closas, and P. M. Djurić, “Indoor tracking: Theory, methods, and technologies,” *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1263–1278, 2015.
- [2] P. Wu, T. Imbiriba, G. LaMountain, J. Vilà-Valls, and P. Closas, “WiFi fingerprinting and tracking using neural networks,” in *Proceedings of the 32nd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2019)*, 2019, pp. 2314–2324.
- [3] S. Niknam, H. S. Dhillon, and J. H. Reed, “Federated learning for wireless communications: Motivation, opportunities, and challenges,” *IEEE Communications Magazine*, vol. 58, no. 6, pp. 46–51, 2020.
- [4] E. Arias-de Reyna, P. Closas, D. Dardari, and P. M. Djurić, “Crowd-based learning of spatial fields for the internet of things: From harvesting of data to inference,” *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 130–139, 2018.
- [5] B. S. Ciftler, A. Albazeer, N. Lasla, and M. Abdallah, “Federated learning for RSS fingerprint-based localization: A privacy-preserving crowdsourcing method,” in *2020 International Wireless Communications and Mobile Computing (IWCMC)*, 2020, pp. 2112–2117.
- [6] F. Yin, Z. Lin, Q. Kong, Y. Xu, D. Li, S. Theodoridis, and S. R. Cui, “Fedloc: Federated learning framework for data-driven cooperative localization and location data processing,” *IEEE Open Journal of Signal Processing*, vol. 1, pp. 187–215, 2020.
- [7] Y. Liu, Y. Kang, C. Xing, T. Chen, and Q. Yang, “A secure federated transfer learning framework,” *IEEE Intelligent Systems*, vol. 35, no. 4, pp. 70–82, 2020.
- [8] T.-M. H. Hsu, H. Qi, and M. Brown, “Measuring the effects of non-identical data distribution for federated visual classification,” *arXiv e-prints*, pp. arXiv–1909, 2019.
- [9] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, “On the convergence of fedavg on non-IID data,” in *Eighth International Conference on Learning Representations (ICLR)*, 2020.
- [10] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, “Scaffold: Stochastic controlled averaging for federated learning,” in *International Conference on Machine Learning*. PMLR, 2020, pp. 5132–5143.
- [11] Q. Li, Y. Diao, Q. Chen, and B. He, “Federated learning on non-IID data silos: An experimental study,” *ArXiv*, vol. abs/2102.02079, 2021.
- [12] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, “Federated optimization in heterogeneous networks,” in *Proceedings of the 1st Adaptive & Multitask Learning Workshop*, Long Beach, California, 2019.
- [13] J. Wang, Q. Liu, H. Liang, G. Joshi, and H. V. Poor, “Tackling the objective inconsistency problem in heterogeneous federated optimization,” in *Advances in Neural Information Processing Systems*, vol. 33, 2020.
- [14] V. Kulkarni, M. Kulkarni, and A. Pant, “Survey of personalization techniques for federated learning,” in *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*. IEEE, 2020, pp. 794–797.
- [15] A. Fallah, A. Mokhtari, and A. Ozdaglar, “Personalized federated learning: A meta-learning approach,” *ArXiv*, vol. abs/2002.07948, 2020.
- [16] S. Ji, T. Saravirta, S. Pan, G. Long, and A. Walid, “Emerging trends in federated learning: From model fusion to federated X learning,” *ArXiv*, vol. abs/2102.12920, 2021.
- [17] Y. Huang, L. Chu, Z. Zhou, L. Wang, J. Liu, J. Pei, and Y. Zhang, “Personalized cross-silo federated learning on non-IID data,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.
- [18] V. Smith, C.-K. Chiang, M. Sanjabi, and A. Talwalkar, “Federated multi-task learning,” in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA, 2017.
- [19] F. Pastor, J. García-González, J. M. Gandarias, D. Medina, P. Closas, A. J. García-Cerezo, and J. M. Gómez-de Gabriel, “Bayesian and Neural Inference on LSTM-Based Object Recognition From Tactile and Kinesthetic Information,” *IEEE Robotics and Automation Letters*, vol. 6, no. 1, pp. 231–238, 2020.
- [20] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, “UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems,” in *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2014, pp. 261–270.