Heterogeneous Zero-Shot Federated Learning with New Classes for On-Device Audio Classification

Gautham Krishna Gudur, Satheesh Kumar Perepu
Ericsson

Deep learning for audio classification is a broad research area with practical applications like Keyword Spotting (KWS), urban sound identification, etc. With the recent compute capabilities vested in resource-constrained devices, there is a natural research focus on audio classification with deep learning on-device. Particularly, on-device Federated learning (FL) is an effective way of extracting insights from different user devices while preserving the privacy of user data [2]. However, new classes with completely unseen data distributions can stream across any device in an FL setting, whose data cannot be accessed by the global server or other users. Moreover, the new class information of one user is not known among other users as well, hence the new classes could be similar or different between users. In addition, there are multiple statistical heterogeneities like label and model heterogeneities across various communication rounds/FL iterations. To this end, we propose a unified zero-shot FL framework to handle these aforementioned challenges in scenarios when new class labels are reported across users with statistical heterogeneities [1].

In order to identify new classes across different users in FL settings, we construct anonymized data without transferring local sensitive data, and identify new classes on this anonymized data. We motivate our framework from the creation of zero-shot *Anonymized Data Impressions* as proposed in [3]. The anonymized feature set $\bar{\mathbf{x}}$ (which has similar properties to original input data) can be synthesized in two steps:

(a) Sample Softmax Values from the Dirichlet distribution [4]. We control the distribution by the *Class Similarity Matrix* which contains important information on how similar the classes are to each other. If the classes are similar, we find similar weights between connections of penultimate layer to the nodes of the classes [3]. We then sample the softmax values from the Dirichlet distribution (with the corresponding concentration parameter which controls its spread).

(b) Creating Anonymized Data Impressions: Once we obtain the softmax values, we compute the synthesized data features (*Data Impressions,* $\bar{\mathbf{x}}$) by minimizing the cross-entropy loss on the model created from randomly initialized input data ($\mathbf{x}$) and the generated sampled softmax values. In this way, the data impressions are created anonymously for each new class without the visibility of original input data.

There are three steps in our proposed FL framework: *Build*, *Local Update* and *Global Update*. The softmax values are sampled with the class similarity matrix in the local devices (local update step), while the anonymized data impressions for new classes are created followed by unsupervised clustering using k-medoids (in the global update step). We also perform parameterized updates [5, 6] to handle statistical heterogeneities. The statistical heterogeneities typically are disparities in models and label distributions with new classes across and within various user devices and federated learning iterations.

We simulate our experiments using *Raspberry Pi 2* with two publicly available datasets on keyword spotting and urban sound classification. We simulate two scenarios for testing our proposed framework – 1) new classes only (homogeneous) with limited users and FL iterations, 2) new classes with statistical heterogeneities in both labels and models with more users and FL iterations, which exhibits near-real-time statistical heterogeneities. The results show effective increase in the local and global update accuracies for both scenarios. Unsupervised clustering with k-medoids on the resultant data impressions for new classes is performed and visualized using PCA, and these new classes are mapped to the respective end-user devices. The new labels are finally added to the overall label set while the corresponding averaged data impressions are added to the public dataset.

References

[1] Gautham Krishna Gudur et al. (2021). Zero-Shot Federated Learning with New Classes for Audio Classification. In *Proc. Interspeech 2021.*

[2] H Brendan McMahan et al. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS 2017).*

[3] Gaurav Kumar Nayak et al. (2019). Zero-Shot Knowledge Distillation in Deep Networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML 2019).*

[4] Thomas Minka (2000). Estimating a Dirichlet distribution.

[5] Gautham Krishna Gudur et al. (2020). Resource-Constrained Federated Learning with Heterogeneous Labels and Models. In *arXiv preprint arXiv:2011.03206.*

[6] Gautham Krishna Gudur et al. (2020) Resource-Constrained Federated Learning with Heterogeneous Labels and Models for Human Activity Recognition. In *Deep Learning for Human Activity Recognition: Second International Workshop, Held in Conjunction with IJCAI-PRICAI 2020.*