Federated Unsupervised Representation Learning using Variational Auto-Encoders

WiML-NeurIPS 2024

Nazreen Shah, Prachi Goyal, Ranjitha Prasad

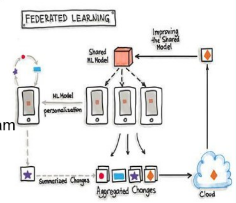
nazreens@iiitd.ac.in

IIIT Delhi

Intellicom lab

Abstract

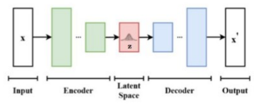
- **Federated Learning (FL)**: Revolutionized machine learning and data privacy.
- **Challenge**: Handling unannotated/unlabeled data.
- **Solution**: Unsupervised FL.
- **Representation Learning**:
  - A popular approach to unsupervised learning.
  - Crucial for generating rich representations for downstream tasks.
- **Variational Autoencoders (VAEs)**:
  - Effective for extracting meaningful representations.
  - Enable uncertainty quantification and reduce overfitting.



Research Question:
Can we formulate a distributed VAE model to achieve federated representation learning?

Contributions:

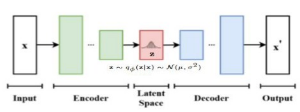
- **IS-FedVAE**: Importance-sampling based Federated Variational Autoencoder framework.
- **Novel Framework**:
  - Based on distributed evidence lower bound for VAE.
  - Enables generating globally relevant samples at the clients.
  - Handles federated learning attributes such as statistical heterogeneity, local epochs, and client participation.
  - Effectiveness of the representations are demonstrated for classification as a downstream task.



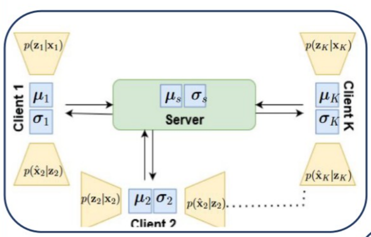
Method

Preliminaries: VAE Decoder-Encoder Parameters are inferred through Evidence Lower Bound (ELBO) over N samples. For a given sample x_i.

$$\log p_\theta(\mathbf{x}_i) \geq \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x}_i)}[\log p_\theta(\mathbf{x}_i|\mathbf{z})] - \mathrm{KL}(q_\phi(\mathbf{z}|\mathbf{x}_i)||p(\mathbf{z})) = \mathcal{L}(\phi, \theta; \mathbf{x}_i).$$



Architecture:



- ➢ Data is distributed across clients.
- ➢ Each of the clients have access to its own local VAE models.
- ➢ Assumption: Global latent distribution satisfies mean field decomposition.
- ➢ Then the optimal global variational distribution is:

$$q_\phi(\mathbf{z}|\mathbf{x}_i) \sim \mathcal{N}(\mu_{s,i}, \sigma^2_{s,i}),$$

$$\mu_{s,i} = \frac{1}{K}\sum_{k=1}^{K} \mu_{k,i}, \quad \sigma^2_{s,i} = \frac{1}{K}\sum_{k=1}^{K}[\sigma^2_{k,i} + \mu^2_{k,i} - \mu^2_{s,i}]$$

Intuition:

- ~~General ELBO:~~

$$\log p_\theta(\mathbf{x}_i) \geq \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x}_i)}[\log p_\theta(\mathbf{x}_i|\mathbf{z})] - \mathrm{KL}(q_\phi(\mathbf{x}_i|\mathbf{z})||p(\mathbf{z})) = \mathcal{L}(\phi, \theta; \mathbf{x}_i).$$

We rewrite the reconstruction error using the importance sampling approach at each client as,

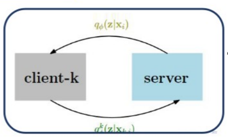
$$\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x}_{k,i})}[\log p^k_\theta(\mathbf{x}_{k,i}|\mathbf{z})] = \mathbb{E}_{q^k_\phi(\mathbf{z}|\mathbf{x}_{k,i})}\left[\frac{q_\phi(\mathbf{z}|\mathbf{x}_i)}{q^k_\phi(\mathbf{z}|\mathbf{x}_{k,i})}\log p^k_\theta(\mathbf{x}_{k,i}|\mathbf{z})\right].$$

KL divergence is computed as follows:

$$\mathrm{KL}(q_\phi(\mathbf{z}|\mathbf{x}_i)||p(\mathbf{z})) = \frac{1}{2}\sum_{i=1}^{d}\left[-\log\left(\sigma^2_{s,i}\right) + \sigma^2_{s,i} + \mu^2_{s,i} - 1\right].$$

Assuming a common Gaussian isotropic prior

➔ The clients communicate the local variational distributions.



➔ The server communicates the global variational distribution.

★ The importance weight introduced in the proposed ELBO, reduces the bias caused by sampling from local variational distributions at the clients.

Scan the QR to read the published work.

Algorithm

Algorithm: IS-FedVAE: Importance Sampling based Federated Variational Autoencoder

Input: Dataset $\mathcal{D}_k$ at the k-th client, Number of communication rounds C, Number of local epochs E, Learning rate η, Initialize $p(\mathbf{z})$ at all clients;

for *C communication rounds* **do**

  At Server:

  Combine all the $q^k_\phi(\mathbf{z}|\mathbf{x}_{k,i})$;

  Communicate $q_\phi(\mathbf{z}|\mathbf{x}_k)$: to compute KLD and importance weights;

  At Client:

  for *E epochs* **do**

   $\forall k \in [K]$, sample mini-batch $B_k \subset \mathcal{D}_k$

   Optimize at each client to obtain $\{\theta_k, \phi_k\}$ by computing the loss based on importance-sampling based reconstruction loss along with KLD ;

   Communicate $q^k_\phi(\mathbf{z}|\mathbf{x}_{k,i}), \forall i$ to the server;

  end

end

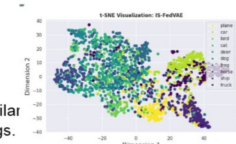
Output: Per-client VAE: $\{\theta_k, \phi_k\}$ after C rounds.

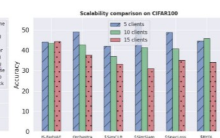
Results

- ➢ Main attributes of FL: Local epochs, Statistical heterogeneity, and client participation.
- ➢ Comparison with baselines on the classification task.
- ➢ Evaluation - Linear probe.
- ➢ Metric - Accuracy.
- ➢ Datasets : CIFAR10, CIFAR100
- ➢ Data partitioning scheme: Dirichlet partitioning.
- ➢ Notations : E - Local epochs, α - Dirichlet parameter, K - No.of Clients, C - Communication rounds.

~~TSNE and Scalability:~~

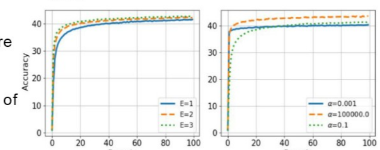
- Representations are well separated.
- Scalable and achieves similar performance across settings.





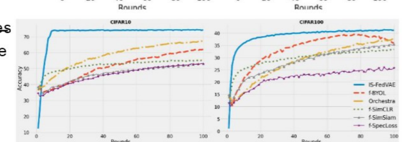
~~Accuracy : E , α~~

- Beneficial in scenarios where there are high communication costs or computational constraints.
- IS-FedVAE is robust to varying levels of statistical heterogeneity.



~~Covergence: Comparison with baselines~~

- Faster convergence - addressing the challenge of communication Bottle-neck



~~Accuracy: Comparison with baselines~~

- Top-1 accuracy (%) comparison under the Linear evaluation protocol for statistically heterogeneous (α = 0.1) setting for CIFAR10 and CIFAR100 datasets.
- The proposed method outperforms the federated baselines.

Methods	CIFAR10	CIFAR100
IS-FedVAE	**77.19**	**43.05**
Orchestra [2]	70.64	35.64
f-BYOL [3]	66.18	38.88
f-SpecLoss [4]	64.53	35.99
f-SimSiam [5]	61.95	36.92
f-SimCLR [6]	58.15	33.49

Conclusions

- ➢ We addressed a crucial problem of unsupervised federated representation learning.
- ➢ Proposed the novel IS-FedVAE using a distributed ELBO formulation.
- ➢ Demonstrated the robustness to FL attributes.
- ➢ Outperformed the state-of-the-art baselines.
- ➢ Future works
  - Other methods to utilize latent distributions in a federated setting.
  - Improving representations to carry out specific downstream tasks.

Acknowledgments & References

I would like to thank Lightmetrics and the SERB-FICCI Prime Minister's Research Fellowship (PMRF) for supporting this work with their generous grant. This work was presented at IEEE ICASSP 2024.

1. https://ml-ops.org/content/three-levels-of-ml-software.html
2. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Artificial intelligence and statistics, pp. 1273–1282, PMLR, 2017
3. Zhang, Xu, et al. "Personalized federated learning via variational bayesian inference." *International Conference on Machine Learning*. PMLR, 2022.
4. E. S. Lubana, C. I. Tang, F. Kawsar, R. P. Dick, and A. Mathur, "Orchestra: Unsupervised federated learning via globally consistent clustering," arXiv preprint arXiv:2205.11506, 2022.
5. J.-B. Grill, F. Strub, F. Altch´e, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar, et al., "Bootstrap your own latent-a new approach to self-supervised learning," Advances in neural information processing systems, vol. 33, pp. 21271–21284, 2020.
6. J. Z. HaoChen, C. Wei, A. Gaidon, and T. Ma, "Provable guarantees for self-supervised deep learning with spectral contrastive loss," Advances in Neural Information Processing Systems, vol. 34, pp. 5000–5011, 2021.J. Z. HaoChen, C. Wei, A. Gaidon, and T. Ma, "Provable guarantees for self-supervised deep learning with spectral contrastive loss," Advances in Neural Information Processing Systems, vol. 34, pp. 5000–5011, 2021.
7. X. Chen and K. He, "Exploring simple siamese representation learning," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 15750–15758, 2021.
8. T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in International conference on machine learning, pp. 1597–1607, PMLR, 2020.