Implementing Scaffold Algorithm in Flower

Miroslav Macko

ABSTRACT

Federated Learning (FL) algorithms face the problem that real-world data are heterogeneous which means that clients have local data that don't conform to the overall data distributions. FL algorithms perform worse when the data is heterogeneous. In this work, we will try to recreate the results of the Stochastic Controlled Averaging for Federated Learning (Scaffold) algorithm as presented by Karimireddy et al.. Scaffold is an FL algorithm with mathematical guarantees of convergence in heterogeneous settings.

Keywords: Scaffold, Federated Learning, Flower

DATA HETEROGENEITY

Kairouz et al. describe that one of the main fundamental challenges of FL is the presence of non-IID data. Kairouz et al. then describe many ways in which data can deviate from being IID:

- **Feature distribution skew** - the same labelled object can have different features. For example, in a handwriting recognition domain, users who write the same words might still have different stroke widths, slants, etc.
- **Label distribution skew** - some clients may see labels in different distribution than the overall distribution. For example, when clients are tied to particular geo-regions, the distribution of labels varies across clients—kangaroos are only in Australia or zoos;
- **Unbalancedness** - a client can hold different amounts of data.
- ...

For further details, please refer to Kairouz et al. (2019) section 3.1 Non-IID Data in Federated Learning.

SCAFFOLD

Karimireddy et al. proposes a new algorithm Scaffold which tries to correct client-drift of updates which is present in FedAvg in non-IID settings. Karimireddy et al. then explain that, intuitively, Scaffold estimates the up-date direction for the server model (c) and the update direction for each client c_i The difference $(c - c_i)$ is then an estimate of the client-drift which is used to correct the local update.

EVALUATION

This evaluation aimed to re-create some of the results visible from table 3 in Karimireddy et al. (2019).

Due to computational limitations, 2% of clients were sampled each round instead of 20%. Sampling 20% clients might be problematic in real-world applications and so using 2% of clients might be better from this perspective.

Instead of s% similarity, concentration was used which relates to the similarity of data. The concentration parameter controls the identicalness among clients. With concentration approaching infinity, all clients have identical distributions to the prior; with concentration approaching zero, on the other extreme, each client holds examples from only one class chosen at random Hsu et al. (2019).

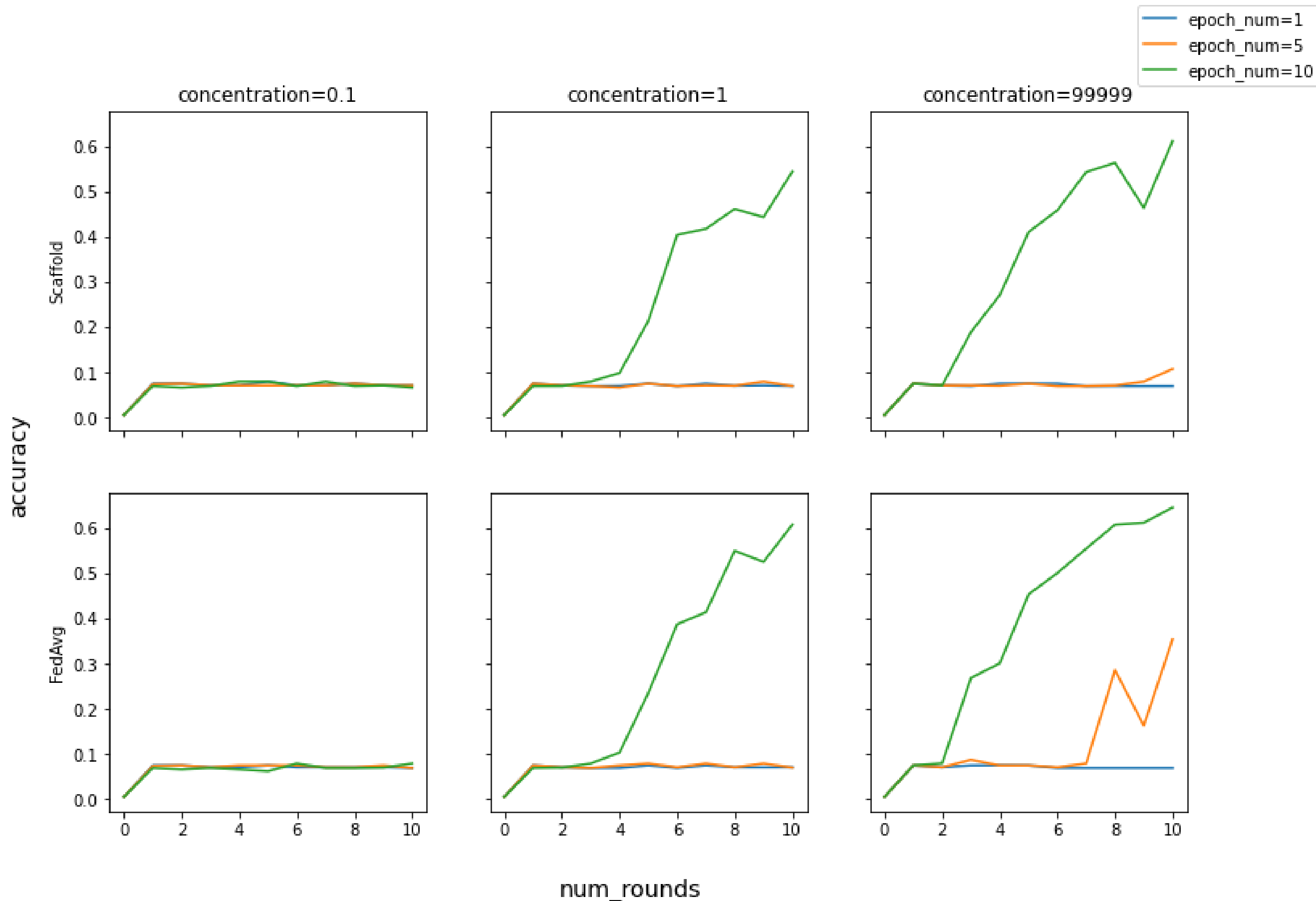
We can see in figure 1 that I haven't managed to replicate what Karimireddy et al. presents. Based on my results, Scaffold and FedAvg algorithms converge at similar rates to better accuracy. This is very different to what Karimireddy et al. presents. There are many potential explanations for this:

- **Different Client learning method** - I wasn't able to find what NN architecture were they using.
- **Different Parameters** - especially the proportion of clients.
- **Potential Mistake** - In my implementation or in the Karimireddy et al. (2019) paper.

IMPLEMENTATION

My implementation can be found in a Google Colab notebook.

Figure 1. Scaffold vs FedAvg - accuracies over rounds comparison



CONCLUSION

I was not able to confirm the results of Karimireddy et al.. Further work is required to do so. The most essential step would be to verify the correctness of a Scaffold implementation. Unfortunately, I haven't found Scaffold implementation online. Once there is a way to verify Scaffold implementation, we will be able to replicate or challenge Karimireddy et al. (2019) results.

ACKNOWLEDGMENTS

Special thanks to professor Nicholas Lane who is in charge of the FL class at the University of Cambridge and to Alexandru-Andrei Iacob, and Lorenzo Sani who were organising practical FL sessions and whose code I used in my implementation.

REFERENCES

Hsu, T.-M. H., Qi, H., and Brown, M. (2019). Measuring the effects of non-identical data distribution for federated visual classification.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G., Eichner, H., Rouayheb, S. E., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., Gruteser, M., Harchaoui, Z., He, C., He, L., Huo, Z., Hutchinson, B., Hsu, J., Jaggi, M., Javidi, T., Joshi, G., Khodak, M., Konečný, J., Korolova, A., Koushanfar, F., Koyejo, S., Lepoint, T., Liu, Y., Mittal, P., Mohri, M., Nock, R., Özgür, A., Pagh, R., Raykova, M., Qi, H., Ramage, D., Raskar, R., Song, D., Song, W., Stich, S. U., Sun, Z., Suresh, A. T., Tramèr, F., Vepakomma, P., Wang, J., Xiong, L., Xu, Z., Yang, Q., Yu, F. X., Yu, H., and Zhao, S. (2019). Advances and open problems in federated learning.

Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S. J., Stich, S. U., and Suresh, A. T. (2019). Scaffold: Stochastic controlled averaging for federated learning.