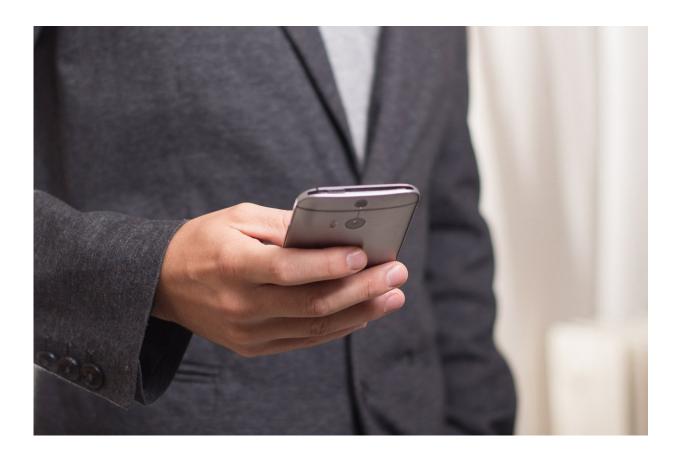# Technique smooths path for AI training in wireless devices

February 1 2022, by Matt Shipman

Credit: CC0 Public Domain

Federated learning is a great tool for training artificial intelligence (AI) systems while protecting data privacy, but the amount of data traffic involved has made it unwieldy for systems that include wireless devices.

A new technique uses compression to drastically reduce the size of data transmissions, creating additional opportunities for AI training on wireless technologies.

Federated learning is a form of machine learning involving multiple devices, called clients. Each of the clients is trained using different data and develops its own model for performing a specific task. The clients then send their models to a centralized server. The centralized server draws on each of those models to create a hybrid model, which performs better than any of the other models on their own. The central server then sends this hybrid model back to each of the clients. The entire process is then repeated, with each iteration leading to model updates that ultimately improve the system's performance.

"One of the advantages of federated learning is that it can allow the overall AI system to improve its performance without compromising the privacy of the data being used to train the system," says Chau-Wai Wong, co-author of a paper on the new technique and an assistant professor of electrical and computer engineering at North Carolina State University. "For example, you could draw on privileged patient data from multiple hospitals in order to improve diagnostic AI tools, without the hospitals having access to data on each other's patients."

There are many tasks that could be improved by drawing on data stored on people's personal devices, such as smartphones. And federated learning would be a way to make use of that data without compromising anyone's privacy. However, there's a stumbling block: federated learning requires a lot of communication between the clients and the central server during training, as they send model updates back and forth. In areas where there is limited bandwidth, or where there is a significant amount of data traffic, the communication between clients and the centralized server can clog wireless connections, making the process slow.

"We were trying to think of a way to expedite wireless communication for federated learning, and drew inspiration from the decades of work that has been done on video compression to develop an improved way of compressing data," Wong says.

Specifically, the researchers developed a technique that allows the clients to compress data into much smaller packets. The packets are condensed before being sent, and then reconstructed by the centralized server. The process is made possible by a series of algorithms developed by the research team. Using the technique, the researchers were able to condense the amount of wireless data shipped from the clients by as much as 99%. Data sent from the server to the clients is not compressed.

"Our technique makes federated learning viable for wireless devices where there is limited available bandwidth," says Kai Yue, lead author of the paper and a Ph.D. student at NC State. "For example, it could be used to improve the performance of many AI programs that interface with users, such as voice-activated virtual assistants."

The paper, "Communication-Efficient Federated Learning via Predictive Coding," is published in the *IEEE Journal of Selected Topics in Signal Processing*. The paper was co-authored by Huaiyu Dai, a professor of electrical and computer engineering at NC State; and by Richeng Jin, a postdoctoral researcher at NC State.

  More information: Kai Yue et al, Communication-Efficient Federated Learning via Predictive Coding, *IEEE Journal of Selected Topics in Signal Processing* (2022). DOI: 10.1109/JSTSP.2022.3142678

Provided by North Carolina State University