MedicalXpress

Training enormous AI models in health care while protecting data privacy

September 7 2021

The new LEARNER platform trains AI models with sensitive health care data by only sharing the internal weightings and workings of the algorithm, keeping all of the data inherently secure in the user's database. Credit: Duke University

Researchers at Duke University and the University of Pittsburgh have developed a platform that allows multiple hospitals and research centers to share private patient data securely to better train machine learning models. The technology could provide single institutions access to advanced predictive tools they could never develop on their own to both advance research and improve patient outcomes.

Called "LEARNER," researchers summarized the platform's development at the National Science Foundation's Convergence Accelerator Expo 2021, an event that shares the program's research portfolio in an exhibition format, like a big science fair.

"AI has incredible potential to improve health data analysis and diagnosis, but it requires a vast amount of data to reach a standard that is acceptable to use in real-life decisions," said Helen Li, the Clare Boothe Luce Professor of Electrical and Computer Engineering at Duke. "And whenever you talk about health care data, there's always a high level of privacy concerns. LEARNER allows health data from many sources to be used to train an AI model without actually sharing any of the sensitive data."

When a machine learning algorithm is trained, it compares the decisions it arrives at to the correct answers, attempts to tweak its inner workings to fix the errors, and repeats the process over and over again until it is no longer improving. These tweaks to its inner workings are referred to as weight parameters.

LEARNER is based on a concept called "Federated learning." In this setup, a single AI model is housed in a central cloud that is provided to users in multiple locations. Each location runs the AI model with its own data and produces a new set of weight parameters, which is in turn sent back to the cloud. The central AI model then uses all of the new weight parameters to develop a single updated algorithm. The process is repeated until the AI model is as good as it can get.

Because only the weight parameters and not the actual data is being shared with the cloud, this technique sidesteps any concerns about data privacy, but the final trained model still represents data from all the entities involved.

"The original information remains hidden on local computers," explained Li. "For a large model, the process typically requires about 50-100 rounds of training between the local entities and the cloud, which sounds like it might take a long time, but in fact only takes a matter of hours."

Built in collaboration with Heng Huang, the John A. Jurenko Endowed Professor at the University of Pittsburgh, the LEARNER prototype demonstrated its usefulness in single-cell multi-omic data and electronic health records. In the former, researchers showed that LEARNER could use scRNA sequencing data to predict the protein markers for associating mRNA sequences with protein production. In the latter, they were able to use health record data to predict the probability of heart failure patients being readmitted within 30 days of being released.

But if all goes according to plan, that's only the beginning. The researchers are developing a user-friendly interface to encourage researchers and doctors to use the platform. Not only would this help LEARNER develop new and better AI health models, the platform also would eventually provide users with hundreds, if not thousands, of pre-trained AI models that they could use in their own laboratories and hospitals.

"We hope LEARNER will be a platform for health experts who want to take advantage of AI but maybe don't know a lot about AI themselves," said Li. "We also hope it will help AI researchers who want to delve into health care and biomedical fields collaborate with one another on large-scale projects."

Li and her colleagues are in talks with a North Carolina-based AI company to continue to develop and potentially commercialize the LEARNER platform.

Provided by Duke University