Google Cloud

ML Powers Clinical & Operational Analytics

Machine learning (ML) has already proved its mettle when working with medical images. For example, when analyzing retinal images, ML can identify diabetic retinopathy better than humans can. Recent research also shows that ML models can look at a retina image and identify a patient's cardiovascular risk – a task that is not humanly possible.

"The retina does have pixels of the patient's blood vessels in the eye. So, there's a hypothesis that it is possible [to assess cardiovascular risk] but humans can't see it. A ML model trained over many examples can easily see that," said Alvin Rajkomar, MD, a Senior Research Scientist at Google.

While these ML use cases are impressive, some clinicians are still waiting for machine learning to have an impact on their daily lives.

> "All of these successes are super cool. But I'm a hospitalist. So, I take care of internal medicine patients, and I don't work with images, for the most part. But there are a lot of tasks that could help doctors like me take care of patients," Rajkomar said.

Indeed, many clinicians would benefit if ML could help them answer questions such as:

- How will my patient progress with a new diagnosis?
- What is my patient's risk of readmission?
- How long will my patient be in the hospital?
- What diagnosis might my patient receive?

Harnessing EHR data

To make such predictions, ML models need to ingest EHR data – something that has only recently become viable. "This actually was not a feasible task before because we didn't have the data. But now the data's there. It's waiting for us to make use of it," Rajkomar said.

It also is now possible to use EHR data in ML models because "we have modern algorithms and software infrastructure to actually harness all that data. At Google, we commonly use TensorFlow, but that software didn't exist 10 years ago," Rajkomar said. TensorFlow is an open-source framework for high performance numerical computation.

Indeed, there's plenty of potential in EHR data, according to Patrik Sundberg, a Google Software Engineer. "There can be a lot of different data points in the electronic health records, especially with long, complicated hospital stays. There's usually a ton of data.

Google Cloud

The classic logistic model might use an average of something like 27 variables. Machine-learning models can use 150,000 variables or more to create more accurate predictions," Sundberg.

The problem is that EHR data is difficult to work with. For every prediction, "you have to first define which variables you actually want to put into the model," Rajkomar said. "Once you define these concepts, extracting them from medical records is extremely hard. Writing SQL queries to extract the right variables in one health center's data might be fine. But if you take a different health center's data, and try to run the same SQL queries, it won't work. And every time you want to address a new prediction, you're stuck on that same cycle — defining the variables, extracting the variables, cleaning them, building a new model."

Scalable models

The challenge is to build models that are scalable and can accommodate more than a single health center's data. To accomplish this ML models need to:

Accept data in a standard format. When working with EHR data, "having a standard format is really a critical ingredient to making sure that you're not creating custom converters for every single piece of data that goes into your model," Rajkomar said.

Fast Healthcare Interoperability Resources (FHIR) makes it possible to arrive at a common representation of data. With data, "you have things like labs and observations" that are "analogous to the pixels in the imaging world, where you have edges and colors," Sundberg said. When using FHIR, all of the healthcare data is mapped into a single format, similar to how DICOM standardizes imaging information.

Accommodate different data volumes

> "It is important to accommodate the fact that some patients actually have decades of data, and some patients only have a few minutes of data when they arrive in an emergency room," Rajkomar said.

Account for time. "If you get Ceftriaxone, Ciprofloxacin, and Flagyl all at one time, it means something very different than if you get those antibiotics spaced out over time," Rajkomar said. The data input representation needs to encode the time element, as it is an important piece of temporal information that leads to accurate predictions.

Include clinical documentation. Critical patient information is often found in the free-form clinical notes. As such, it's important to include these notes, not just structured information in ML.

By taking these steps, healthcare organizations can use EHR data in ML models – making it possible to finally leverage the vast knowledge that exists in these records to improve patient care.

For more information visit the healthcare website:
https://cloud.google.com/solutions/healthcare/