AWS Government, Education, & Nonprofits Blog

An Eye on Science: How Stanford Students Turned Classwork into Their Life's Work

on 04 OCT 2016 | in Education* | Permalink | Share

Imagine visiting your doctor only to discover that you are losing your vision. That's the reality for millions of people with diabetes every year around the world. Diabetic retinopathy causes progressive damage to the retina and is one of the leading causes of blindness. It afflicts more than 100 million people. Early intervention helps in 90% of the cases, but often the diagnosis happens too late due to the cost, complexity, and expertise required for the scanning exam.

Early last year, Stanford University students, Jason Su and Apaar Sadhwani, took the Project in Mining Massive Data Sets course taught by Dr. Anand Rajaraman and Dr. Jeffrey Ullman. The course gives students practical experience in data mining and machine learning algorithms for analyzing large amounts of data. Students undertake team projects of their own design with the mentorship of professors and the cloud computing power of Amazon Web Services (AWS). AWS provided platform credits to the students and instructor as part of the curriculum.

Building on their project that used AWS in the classroom, Jason and Apaar dared to ask the question: can deep learning be applied to quickly and cost-effectively identify the precursors to blindness caused by diabetic retinopathy?

Deep learning is like machine learning; it allows the computer to adjust and fine-tune a general model based on the data it sees and "learns" from it, rather than having a human enter parameters or design complicated rules for the machine to follow. In the past, algorithms to understand retinal images were all based on narrow approaches assuming ideal images. They couldn't manage the many common artifacts in pictures and they required a prohibitive amount of compute power, making them impractical for a real clinic. In practice, doctors painstakingly scan through eye images looking for tiny lesions that suggest the onset of the disease. This process of finding needles in a haystack takes several minutes for each image. Deep learning could be the key to automating this process.

Based on their experiences in class, Jason and Apaar knew AWS provided the compute and storage resources for an automated diagnostic solution based on eye imagery. This was the starting point for their journey from "grass roots to [a] sophisticated infrastructure" for predicting eye diseases.

Using Amazon Elastic Compute Cloud (Amazon EC2), Amazon Simple Storage Service (Amazon S3), and Amazon Elastic Block Store (Amazon EBS), Apaar and Jason have imported more than 80,000 eye images and they are trying to access tens of thousands more from other sources, both local to Stanford and international. They are using AWS to manage, crunch, and review the many gigabytes of data.