Google AI precision medicine tool open source

Google has made available to the open source community its new deep-learning based technology that will be useful in broadening our knowledge of genomics. DeepVariant, developed by the Google Brain team, is designed to reconstruct the true genome sequence from high-throughput sequencing (HTS) sequencer data with significantly greater accuracy than previous classical methods.

Essentially, DeepVariant is there to transform the task of genome reconstruction into an image classification problem which fits better to deep learning technology. Transforming data to images is one of the techniques commonly employed by deep learning practitioners.

"We started with GIAB [Genome in a Bottle] reference genomes, for which there is high-quality ground truth (or the closest approximation currently possible). Using multiple replicates of these genomes, we produced tens of millions of training examples in the form of multi-channel tensors encoding the HTS instrument data, and then trained a TensorFlow-based image classification model to identify the true genome sequence from the experimental data produced by the instruments," explain Mark DePristo and Ryan Poplin, of the Google Brain team, in a blog post.

It is hoped that the release of DeepVariant as an open source software will encourage collaboration, thus accelerating the use of this technology to solve real world problems. In line with this, the Google Cloud Platform (GCP) is also being tapped to help run the software on other complex, scientific and health research projects. DeepVariant workflows may be deployed on GCP in configurations optimised for low-cost and fast turnarounds using scalable GCP technologies like the Pipelines API. This paired set of releases provides a smooth ramp for users to explore and evaluate the capabilities of DeepVariant in their current compute environment while providing a scalable, cloud-based solution to satisfy the needs of even the largest genomics, according to the blog post.

"DeepVariant is the first of what we hope will be many contributions that leverage Google's computing infrastructure and ML [machine learning] expertise to both better understand the genome and to provide deep learning-based genomics tools to the community," say DePristo and Poplin. "This is all part of a broader goal to apply Google technologies to healthcare and other scientific applications, and to make the results of these efforts broadly accessible."

Source: Google Research Blog
Image Credit: Pixabay

Published on : Tue, 12 Dec 2017