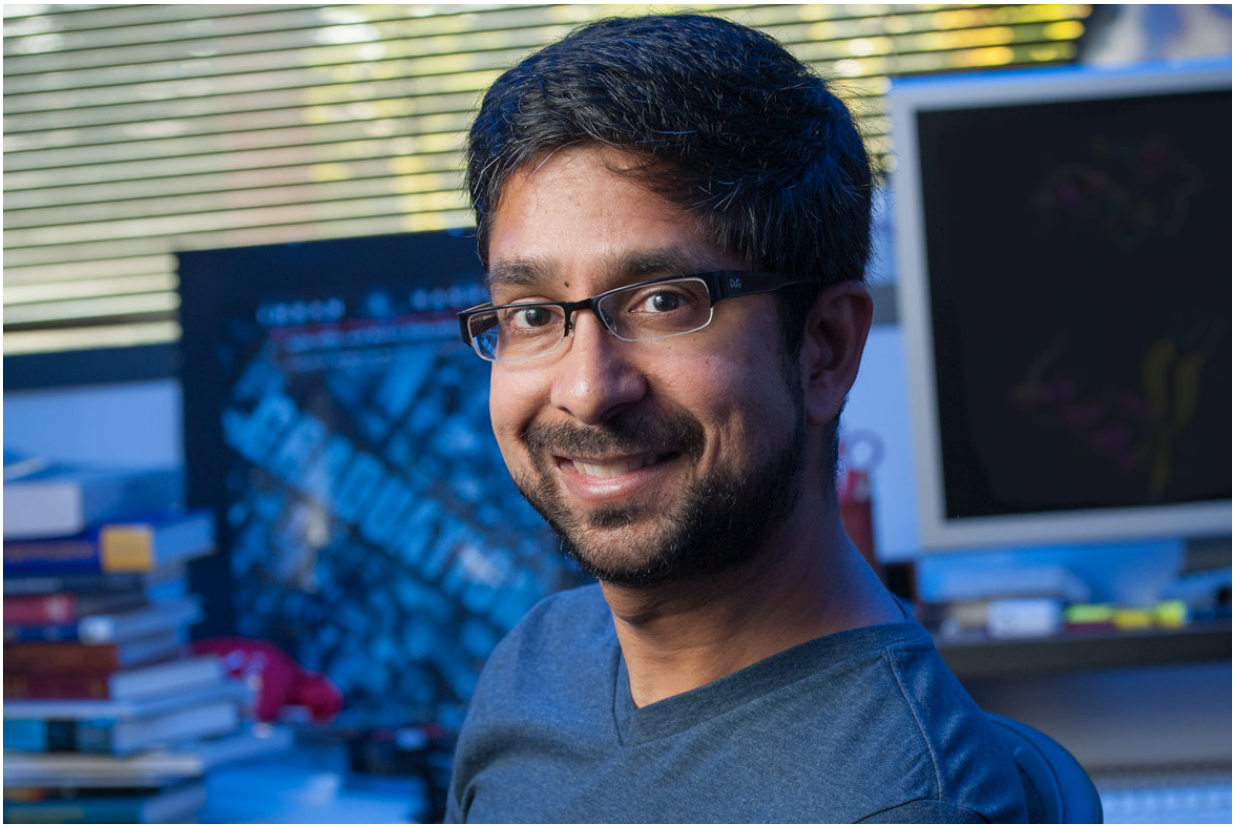
Researchers create deep learning algorithm that could boost drug development

April 3 2017, by Taylor Kubota



Stanford chemistry Professor Vijay Pande and his students see a future for machine learning in the early stages of drug development. Credit: L.A. Cicero

Artificially intelligent algorithms can learn to identify amazingly subtle information, enabling them to distinguish between people in photos or to

screen medical images as well as a doctor. But in most cases their ability to perform such feats relies on training that involves thousands to trillions of data points. This means artificial intelligence doesn't work all that well in situations where there is very little data, such as drug development.

Vijay Pande, professor of chemistry at Stanford University, and his students thought that a fairly new kind of deep learning, called one-shot learning, that requires only a small number of data points might be a solution to that low-data problem.

"We're trying to use machine learning, especially deep learning, for the early stage of drug design," said Pande. "The issue is, once you have thousands of examples in drug design, you probably already have a successful drug."

The group admitted the idea of applying one-shot learning to drug design problems was farfetched—the data was likely too limited. However, they'd had success in the past with machine learning methods requiring only hundreds of data points, and they had data available to test the one-shot approach. It seemed worth a try.

Much to their surprise, their results, published April 3 in *ACS Central Science*, show that one-shot learning methods have potential as a helpful tool for drug development and other areas of chemistry research.

Moving from images to molecules

Other researchers have successfully applied one-shot learning to image recognition and genomics, but applying it to problems relevant to drug development is a bit different. Whereas pixels and bases are fairly natural types of data to feed into an algorithm, properties of small molecules aren't.

To make molecular information more digestible, the researchers first represented each molecule in terms of the connections between atoms (what a mathematician would call a graph). This step highlighted intrinsic properties of the chemical in a form that an algorithm could process.

With these graphical representations, the group trained an algorithm on two different datasets - one with information about the toxicity of different chemicals and another that detailed side effects of approved medicines. From the first dataset, they trained the algorithm on six chemicals and had it make predictions about the toxicity of the other three. Using the second dataset, they trained it to associate drugs with side effects in 21 tasks, testing it on six more.

In both cases, the algorithm was better able to predict toxicity or side effects than would have been possible by chance.

"We worked on some prototype algorithms and found that, given a few data points, they were able to make predictions that were pretty accurate," said Bharath Ramsundar, who is a graduate student in the Pande lab and co-lead author of the study.

However, Ramsundar cautioned that this isn't a "magical" technique. It was built off of several recent advances in a particular style of one-shot learning and it works by relying on the closeness of different molecules, as indirectly indicated by their formula. For example, when the researchers trained their algorithm on the toxicity data and tested it on the side effect data, the algorithm completely collapsed.

An experimentalist's help

People concerned about AI taking jobs from humans have nothing to fear from this work. The researchers envision this as groundwork for a

potential tool for chemists who are early in their research and trying to choose which molecule to pursue from a set of promising candidates.

"Right now, people make this kind of choice by hunch," Ramsundar said. "This might be a nice compliment to that: an experimentalist's helper."

Beyond giving insight into drug design, this tool would be broadly applicable to molecular chemistry. Already, the Pande lab is testing these methods on different chemical compositions for solar cells. They have also made all of the code they used for the experiment open source, available as part of the DeepChem library.

"This paper is the first time that one-shot has been applied to this space and it's exciting to see the field of machine learning move so quickly," Pande said. "This is not the end of this journey - it's the beginning."

Provided by Stanford University

Citation: Researchers create deep learning algorithm that could boost drug development (2017, April 3) retrieved 17 April 2024 from https://techxplore.com/news/2017-04-deep-algorithm-boost-drug.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.