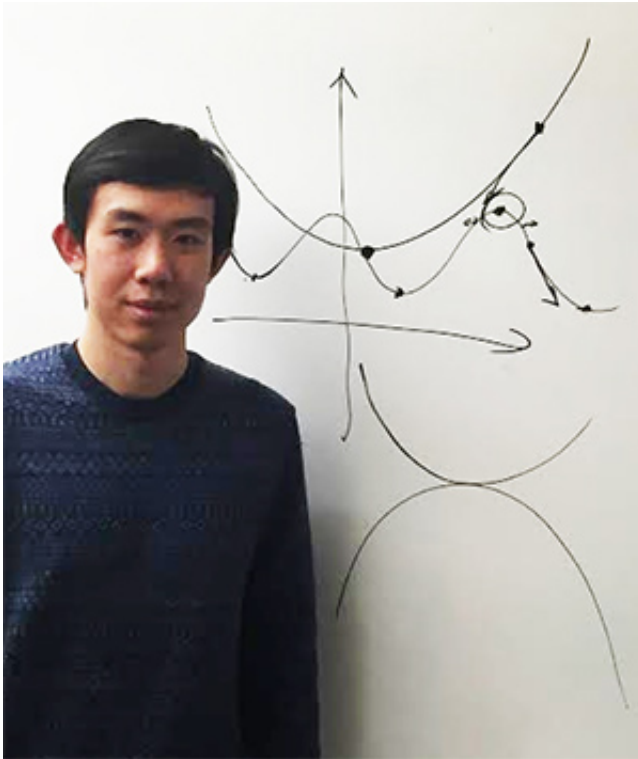


Improving machine learning with an old approach

December 22 2015, by Grayson York



Rong Ge is an assistant professor of computer science.

Computer scientist Rong Ge has an interesting approach to machine learning. While most machine learning specialists will build an algorithm which molds to a specific dataset, Ge builds an algorithm which he can guarantee will perform well across many datasets.

A paper he wrote as a postdoc at Microsoft Research, *Escaping From Saddle Points—Online Stochastic Gradient for Tensor Decomposition*, describes how a programmer can use the imprecision of a common [machine learning algorithm](#), known as stochastic gradient descent, to his advantage.

Normally this algorithm is used to speed up a slow learning process by only approximating the correct solution rather than working harder to get precision; however, Ge and his colleagues found that the small amount of "noise" created by the algorithm can be the saving grace of an algorithm which would otherwise be trapped by its own perfectionism.

"This algorithm is not normally used for this purpose," Ge says, "It is normally used as a heuristic to approximate the solution to a problem."

Noise allows the algorithm to escape from something called a "saddle point" on the function which the stochastic gradient is trying to optimize, which looks sort of like a sine wave. Ge describes gradient descent as being like a ball rolling down a hill. When on the slope of the hill it always seeks the lowest point, but if it is at a saddle point, a high point on a "ridge" between two different slopes, it will not start rolling.

Stochastic gradient descent remedies this problem by jostling the ball enough to start it rolling down one of the slopes. But one cannot be sure which way it will roll.

The results he has obtained relate to a certain branch of machine learning known as unsupervised learning. One common problem from unsupervised learning is "clustering," in which the algorithm attempts to find clusters of points which are similar to each other while different from the other points in the set. The algorithm then labels each of the clusters which it has found, and returns its solution to the programmer.

A key requirement for the final result of the algorithm to be correct is that the two slopes end at low points of equal depth, which represent optimal solutions. Often two solutions will appear different at first glance, but will actually represent the same set of clusters, different only because the labels on clusters were switched. Ge said one of the hardest parts of his process was designing functions that have this property.

These results are guaranteed to hold so long as the dataset is not provided by someone who has specifically engineered it to break the algorithm. If someone has designed such a dataset the problem becomes "NP hard," meaning that there is little hope for even the best algorithms to solve it.

Hopefully we will see more growth in this field, especially interesting results such as this which find that the weaknesses associated with a certain algorithm can actually be strengths under different circumstances.

More information: Escaping From Saddle Points — Online Stochastic Gradient for Tensor Decomposition.
arxiv.org/abs/1503.02101

Provided by Duke University

Citation: Improving machine learning with an old approach (2015, December 22) retrieved 19 April 2024 from <https://phys.org/news/2015-12-machine-approach.html>

| |
|--|
| <p>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.</p> |
|--|