

More is less: Complex computer models can involve thousands of variables

June 2 2010, by Larry Hardesty



Explaining the relationships between observable data (blue dots) can involve complicated mathematics that correlates each data point with each of the others (blue lines). But a "hidden variable" that describes general properties of all the data points (green dot) can make the mathematics much simpler (green lines). Graphic: Christine Daniloff

(PhysOrg.com) -- The architect Mies van der Rohe is famous for promoting the slogan "less is more." But if Venkat Chandrasekaran, a graduate student in the Department of Electrical Engineering and Computer Science, had a slogan for his own work, it might be "more is less."



Science, engineering and other quantitative disciplines are largely concerned with uncovering the mathematical relationships between data points — such as energies of molecules, measurements of temperature or gene activity, or stock prices. In most cases, adding more data points just makes the math more complicated. But sometimes it makes it simpler. And for many types of calculations, if there are additional data points that will make them simpler, Chandrasekaran's techniques will find them.

To see how adding data points can mean simpler calculations, suppose that you're trying to understand the relationships between a bunch of stocks in the same industry sector — say, Apple, Gateway, Dell, Hewlett-Packard and other computer manufacturers. On the one hand, an increase in Apple's share price could mean a decrease in, say, Dell's, because Apple and Dell compete for a limited pool of computer buyers' dollars; on the other hand, if large institutional investors are bullish about computer stocks in general, an increase in Apple's stock could indicate an increase in Dell's as well.

It might be possible to build a complicated mathematical model that, on the basis of considerations like the companies' price-to-earnings ratios, trade volumes and revenues determines whether an increase in Apple's share price will cause an increase or decrease in Dell's — and Gateway's, and Hewlett-Packard's, and so on. But it might also turn out that a single extra variable — say, the average price of all the companies' stock provides a good indication of general trends in the sector. Since the new variable accounts for institutional investors' enthusiasm or skittishness, the relationships between the individual stocks no longer have to. The overall calculation becomes much simpler.

Irrelevant referent

In this case, Chandrasekaran's techniques would tell you only that adding



another variable — the average stock price — simplifies the overall calculation. They wouldn't tell you why. And indeed, the extra variable could turn out to be something more complicated than an average. It might factor in the price-to-earnings ratios of some companies, the revenues of others, the share prices of still others, and so on. A savvy analyst might be able to deduce that this new, more complex variable represents the trading strategies of a bunch of large hedge funds that concentrate on the computer industry. But then again, it could be that no one has any idea what the new variable refers to.

"There's this temptation that I even had initially, that you can sort of discover hidden variables," says Chandrasekaran. "And that's true: You can discover hidden variables. But it's not going to be easy to attribute meaning to these hidden variables." For most purposes, however, that may not matter. "From the mathematical point of view, just putting these things in helps you simplify," Chandrasekaran says. If the added variable helps you predict Dell's <u>share price</u> from Apple's, does it really matter what it refers to — or whether it refers to anything at all?

Getting to the bottom

At the most recent Symposium on System Identification, hosted by the International Federation of Automatic Control, Chandrasekaran and MIT Professors of Electrical Engineering Alan Willsky and Pablo Parrilo described their approach to finding hidden variables that simplify calculations.

Generally, computer science is concerned with questions of computational complexity: Given a particular algorithm, you want to know whether a computer can execute it quickly, slowly or never. So <u>computer science</u> provides some standard methods for calculating the complexity of mathematical models.



If you have an equation that describes the complexity of a <u>mathematical</u> <u>model</u>, you want to find its minimum values: where the complexity is lowest, the model is simplest, and thus easiest to work with. If you imagine the graph of the equation as a complex surface with lots of peaks and troughs, you want to find the bottom of the deepest trough.

But that in itself can be a prohibitively complex process. Computer scientists have developed a host of methods for analyzing such equations and finding solutions that are probably near the bottom of a trough in a particular region of the graph. For certain types of problems, however, the techniques developed by Chandrasekaran and his colleagues are mathematically guaranteed to find the bottom of the graph's lowest trough.

According to Ben Recht, an assistant professor in the University of Wisconsin's computer sciences department, "There are a lot of people who would be surprised if you told them that you could solve this particular hidden-variable problem using [Chandrasekaran's] methods." He adds, however, that "it's not a general-purpose tool, even for these hidden-variable problems." Chandrasekaran agrees. In fact, he prefers to describe his methods as "tricks" rather than "techniques," because it might require some mathematical insight to determine how to apply them in any particular case.

Still, Recht says, "he's shown that in a relatively large set of cases, you can actually use this. And it's a first step to explore the space of what sorts of problems can be solved using this technology."

Provided by Massachusetts Institute of Technology

Citation: More is less: Complex computer models can involve thousands of variables (2010, June 2) retrieved 28 April 2024 from



https://phys.org/news/2010-06-complex-involve-thousands-variables.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.