

Scientists devise new method to solve significant variables conundrum

October 26 2015

Scientists at Columbia University, the University of California, San Diego (UCSD) and Harvard University have presented an alternative method to address the challenge of using significant variables to make useful predictions in areas such as complex disease.

Shaw-Hwa Lo and Tian Zheng of Columbia, Adeline Lo of UCSD and Herman Chernoff of Harvard present findings in a paper to appear in *Proceedings of the National Academy of Sciences* on Monday, October 26, that demonstrates that statistically significant variables are not necessarily predictive. In addition, very predictive variables do not necessarily have to appear significant and thereby evade a researcher using [statistical significance](#) as a criterion to evaluate variables for prediction.

Statistical significance is a traditional, long-standing measure in any researcher's toolbox but thus far, scientists have been puzzled by the inability to use results of statistically significant variants in complex diseases to make predictions useful for personalized medicine. Why aren't significant variables leading to good prediction of outcomes? This conundrum affects both simple and complex data in a broad range of science and social science fields.

In their findings, the authors demonstrate that what makes variables good for prediction versus significance depends on different properties of the underlying distributions. They suggest that progress in prediction requires efforts toward a new research agenda of searching for a novel

criterion to retrieve highly predictive variables rather than highly significant variables.

They also present an alternative approach, the Partition Retention [method](#), which displays strong power in prediction. The researchers applied the method to a well-known [breast cancer](#) dataset, the van't Veer dataset, and reduced the prediction error rate from 30% to 8%, finding breast cancer genes that are highly predictive - and not significant.

Their results show that using their method to examine the top five interacting breast cancer genes they were able to find predicted breast cancer relapse effectively, when the outcome would not have appeared using significance measures. Previous methods were only 70% correct in predicting something as significant as breast cancer relapse. Using the new method and avoiding significance as a criterion, scientists correctly predicted such an outcome with 92% accuracy.

"What we're saying here is that using the previously very well-known methods might not be appropriate when we care about predicting important outcomes," says Professor Lo. "Our alternative approach seems to do very well in [prediction](#), and is relevant for many scientific fields."

More information: Why significant variables aren't automatically good predictors, www.pnas.org/cgi/doi/10.1073/pnas.1518285112

Provided by Columbia University

Citation: Scientists devise new method to solve significant variables conundrum (2015, October 26) retrieved 24 April 2024 from <https://phys.org/news/2015-10-scientists-method-significant-variables-conundrum.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.