

Noisy data facilitates Dartmouth investigation of breast cancer gene expression

January 22 2015

Researchers from Dartmouth's Norris Cotton Cancer Center, led by Casey S. Greene, PhD, reported in *Pacific Symposium on Biocomputing* on the use of denoising autoencoders (DAs) to effectively extract key biological principles from gene expression data and summarize them into constructed features with convenient properties.

"Cancers are very complex," explained Greene. "Our goal is to measure which genes are being expressed, and to what extent they're being expressed, and then automatically summarize what the cancer is doing and how we might control it."

Normally, it is difficult to apply computational models across different studies because the [gene expression data](#) is "noisy," meaning that there are many factors that differ in the way [gene expression](#) is measured. To begin their analysis, Greene's team added more noise to the data and then trained a computer to remove the noise. To remove the noise, the computer had to learn about key underlying features of [breast cancer](#). "This approach of removing noise makes the models we constructed more generally applicable," Greene said.

Greene and the Dartmouth team studied DAs, which train computers directly on the data without requiring researchers to provide known [biological principles](#) to the computer, as a method to identify and extract complex patterns from [genomic data](#). The model that the computer constructs can then be compared to previous discoveries to understand where data supports those discoveries and where the data raises new

questions. The performance of DAs was evaluated by applying them to a large collection of breast cancer gene expression data. Results show that DAs were able to recognize changes in gene expression that corresponded to the cancers' molecular and clinical information.

"These techniques and findings will enable others to use the DAs to evaluate gene expression data in a variety of disease sites," reported Greene. "While noise in data is usually viewed as a problem, adding [noise](#) to data can actually be a good thing because it can help reveal the underlying signal. When we did this to analyze data from breast cancers, we found gene expression features that generalize across studies and represent important clinical factors."

Next for Greene's research team are more complex models that take multiple levels of regulation into account. Their goal is to develop methods that not only model data but that can automatically explain to researchers what the models have learned.

More information: [psb.stanford.edu/psb-online/pr...
edings/psb15/tan.pdf](https://psb.stanford.edu/psb-online/pr...edings/psb15/tan.pdf)

Provided by The Geisel School of Medicine at Dartmouth

Citation: Noisy data facilitates Dartmouth investigation of breast cancer gene expression (2015, January 22) retrieved 27 April 2024 from <https://medicalxpress.com/news/2015-01-noisy-dartmouth-breast-cancer-gene.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>
--