

December 22, 2011

A novel mathematical framework for discovery from comparison of large-scale datasets

A team of scientists, led by a faculty member of the Scientific Computing and Imaging (SCI) Institute at the University of Utah, have developed a novel mathematical framework for the comparison of multiple large-scale datasets arranged in tables of different row dimensions but the same column dimensions.

The number of such datasets, recording different aspects of a single phenomenon, is fast growing in science and medicine. Gaining access to the full information that these datasets store requires mathematical frameworks that can compare and contrast them in order to find the similarities and dissimilarities among them. Until now only one such framework existed, which was limited to a comparison of two datasets at a time.

In an article published in the journal *PLoS One* on December 22, 2011, Orly Alter, USTAR associate professor of bioengineering and human genetics at SCI, her lab alumna, Sri Priya Ponnappalli, in collaboration with Charles F. Van Loan of Cornell University and Michael A. Saunders of Stanford University, formulated a novel generalization of the existing framework that enables comparison of more than just two datasets at a time. The team demonstrated the novel framework in comparative modeling of the cellular activities of three evolutionarily disparate organisms – human and budding and fission yeasts. The mathematical model successfully identified and separated cellular events that are common to the human and yeasts from those that are exclusive to either one of the organisms.

Alter's previous comparative modeling of the cellular activities of just two of the organisms – human and budding yeast – led to the computational prediction of a new mode of biological regulation, which she then experimentally verified in collaboration with John F. X. Diffley of Cancer Research UK. In a recent comparative modeling of the genomes of just two cell types – normal blood and brain cancer cells – she discovered a new link between a brain tumor's genome and a patient's prognosis, which offers insights into the cancer's formation and growth, and suggests promising targets for drug therapy.

Just as these discoveries were made possible by the ability to compare between two datasets, the mathematical framework formulated by Alter and her team, which enables – for the first time – a comparison of more than two datasets at a time, promises to lead to discoveries that would have been impossible without it. Although this mathematical framework was developed with applications in biotechnology in mind, it could similarly be used to make discoveries in any of the many areas where large-scale datasets are being accumulated today.

Funding for Alter's research comes from National Science Foundation (NSF) CAREER Award DMS-0847173, National Human Genome Research Institute (NHGRI) R01 Grant HG-004302 and the Utah Science, Technology and Research (USTAR) initiative.

To read the December 22, 2011 article in *PLoS One* go to:
<http://dx.doi.org/10.1371/journal.pone.0028072>