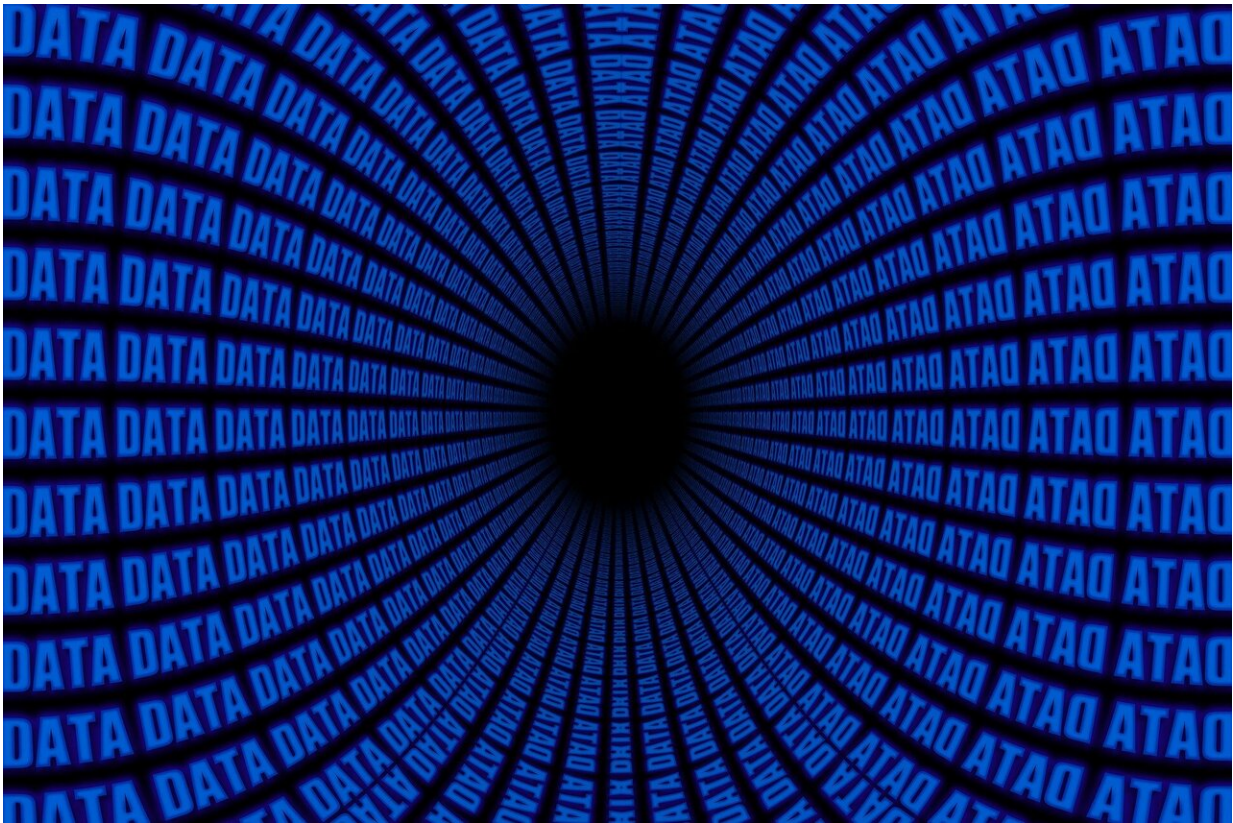


Advancement made in the visualization of large, complex datasets

December 2 2019



Credit: CC0 Public Domain

An improvement to the premier data visualization tool t-distributed Stochastic Neighborhood Embedding (t-SNE), called optimized-t-SNE (opt-SNE), shines new light on researchers' ability to view exactly what

is in their datasets.

opt-SNE is an advancement of the widely used t-SNE created nearly 10 years ago. While t-SNE can accurately analyze approximately half a million [cells](#) in any given sample, in recent years, single cell datasets have become much larger. With opt-SNE, researchers can now visualize data from samples containing tens of millions of cells with unprecedented resolution.

The development of opt-SNE was led by Anna Belkina, MD, Ph.D., assistant professor of pathology and laboratory medicine at Boston University School of Medicine (BUSM).

In addition to its capacity to properly process big datasets, opt-SNE was also able to successfully visualize very small, distinct populations of cells in the [blood samples](#) tested (with each cell in these groups as rare as one in a hundred thousand of the total number of cells in the sample). Prior to opt-SNE, this accurate, large-scale visualization with simultaneous magnification of miniscule populations was not possible. "t-SNE was originally a "one-size-fits-all" algorithm, but opt-SNE computations are tailored to each individual dataset and this allows both a birds-eye and up-close view of what is in your sample. With opt-SNE, both the haystack and the needles within it can be seen," explained Belkina, the corresponding author of the study. "It is a particularly [valuable tool](#) for the investigation of cytometry and single cell transcriptomics data".

opt-SNE allows researchers to pinpoint previously undetectable features that distinguish diseased samples from controls. This new lens into disease states may reveal novel targets for therapies as well as new biological phenomena. This approach is already in use by multiple research groups due to Belkina's ongoing collaborations with developers of major single cell data analysis platforms who enabled opt-SNE implementation into the Omiq.ai cloud analysis platform (Christopher

Ciccolella, MS) and FlowJo software (Josef Spidlen, Ph.D. and Richard Halpert, Ph.D.) and co-authored the manuscript. An [open-source](#) opt-SNE package has also been released.

Additional co-authors of the study, which appears online in *Nature Communications*, include Rina Anno, Ph.D. and Jennifer Snyder-Cappione, Ph.D.

More information: Anna C. Belkina et al, Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets, *Nature Communications* (2019). [DOI: 10.1038/s41467-019-13055-y](https://doi.org/10.1038/s41467-019-13055-y)

Provided by Boston University School of Medicine

Citation: Advancement made in the visualization of large, complex datasets (2019, December 2) retrieved 27 July 2024 from <https://phys.org/news/2019-12-advancement-visualization-large-complex-datasets.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.