From machine learning to human understanding

Daniel Masters*[1]

[1]Caltech / IPAC – United States

Abstract

A common knock on machine learning is that, while the results can be remarkably accurate, the information and relationships underlying the results remain opaque. This is a especially problematic for scientific investigations, where the ultimate goal is understanding. In this talk I will argue that powerful visualization tools based on machine learning-based dimensionality reduction techniques such of SOM, t-SNE, and UMAP are potentially crucial in terms of exploratory analysis and scientific investigation of massive astronomical datasets. Not only can they provide valuable intution for large, multidimensional data distributions, they can be used to fundamentally constrain the distibution of the galaxy population, yielding real insights about galaxy evolution. They can also be used for efficient selection of galaxy subpopulations, identification of problematic data, and more. Such manifold learning-based visualization tools should become a standard part of the analytical toolkit in the era of massive datasets in order to facilitate human understanding of these data.

Keywords: Visualization, Galaxy evolution, Modeling, Selection, Manifold learning

*Speaker