Machine-learn a dust spectrum: what kind of disk is this?

Grant Kennedy

University of Cambridge, UK

The era of big data has clearly arrived, yet our ability to sort through this data is still developing. Many instances arise where we wish to characterize large numbers of targets, yet want to ensure reliable enough results that rare objects can be identified. When the characterization involves fitting of models, one challenge is therefore ensuring that the appropriate model is used.

Most of my work is on circumstellar disks, and the standard way to characterize these systems in the first instance is by fitting stellar and dust models to photometry and spectra over a wide range of wavelengths. When dealing with systems on an individual basis, figuring out which type of model to fit is relatively easy because a trained eye can discern among different possibilities, and usually chose the most appropriate one. For large numbers of systems however, choosing the best model in an automated way can be challenging and time-consuming.

A possible way to aid this challenge is with some form of machine learning. The basic concept is that a set of "training" data with known properties are used to condition a model, which is then applied to real data with unknown properties. My overall aim with this work is to use machine learning to classify photometry and spectra of stars and disks. Initially the goal is to inform automated model fitting, for example to discern among protoplanetary and debris disks. Further work will be to identify specific spectral features, for example silicates, PAHs, Be-star emission lines, and perhaps contamination from extragalactic sources, primarily in Spitzer IRS spectra (of which there are over 20,000).

This work is in the early stages, so my hope is to present it at the Cosmic Dust meeting and get feedback and ideas from dust experts, and discuss if and how it should be developed to make it applicable to a wider audience.