**Raquel A. Romano**

Home
Projects
Talks
Publications
Links
CV

Raquel A. Romano
Lawrence Berkeley National Laboratory
1 Cyclotron Road
50F-1650
Berkeley, CA 02139
(510) 486-5699
romano@hpcrd.lbl.gov

## Current Projects

### Recognition of Supernovae using Support Vector Machines

This work demonstrates the great potential impact that supervised learning has to improve the efficiency of large-scale digital sky surveys that are slated to collect terabytes of nightly imagery in search of celestial objects (SNAP, LSST, DES, Pan-STARRS). The Nearby Supernova Factory (SNfactory) is an international project to obtain spectrophotometry data on a large sample of Type Ia supernovae in a nearby redshift range in order to measure the expansion history of the universe. Each night, the SNfactory receives about 80 GB of wide field CCD imaging data and analyzes them with special-purpose image processing software. Several hundred thousand subimages with potential supernovae must be narrowed down to only a few good supernovae candidates to be visually scanned by humans, who determine which should be sent on for spectroscopic observation.
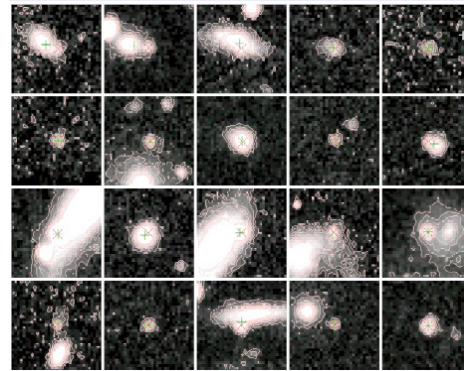


Example images of SNe discovered by the Nearby Supernova Factory. Classification techniques have improved the detection rate while decreasing the false positive rate.

We have used supervised learning techniques (Support Vector Machine (SVMs), boosted decision trees, random forests) to automatically classify all incoming subimages on a nightly basis and rank-order them by the classifier decision value, allowing astrophysicsts to quickly examine the 20 or so most promising candidates arriving each morning. The difficulties of noisy and corrupt imagery resulting in high levels of feature uncertainty and extremely imbalanced and overlapping positive and negative data sets make this a challenging learning problem. High accuracy is achieved by preprocessing features to transform peaked, skewed distributions to more Gaussian densities, and by oversampling positives an selectively undersampling negatives and iteratively training multiple classifiers for improved supernova recognition on unseen test data.

Supernova Recognition using Support Vector Machines., R. Romano, C. Aragon, and Chris Ding. *Proceedings of the 5th International Conference of Machine Learning Applications.* December 14-16, 2006.

Supernova Recognition Using Support Vector Machines, Neyman Seminar, Department of Statistics, UC Berkeley, September 20, 2006.

How To Find More Supernovae With Less Work, S. Bailey, et. al., AAS 209, Winter 2007.

### Dimensionality Reduction of Observed and Synthetic Supernova Spectra

While spectra taken from observed Type Ia supernovae (SNe Ia) exhibit remarkably similar shapes, there is an intrinsic diversity in particular spectral features among them. Identifying the wavelength regions that characterize the spectral variability within and across different subtypes of SNe Ia would be useful for improving cosmological parameter estimation. Classical dimensionality reduction methods such as PCA have previously been used to decompose spectra into components maximally varying components (see James, et.al., Spectral diversity of Type Ia Supernovae). We experiment with alternative decompositions such as non-negative matrix factorization (NMF) and nonlinear dimensionality reduction methods with the goal of find decompositions with the following properties:



NMF components

NMF reconstruction

PCA reconstruction
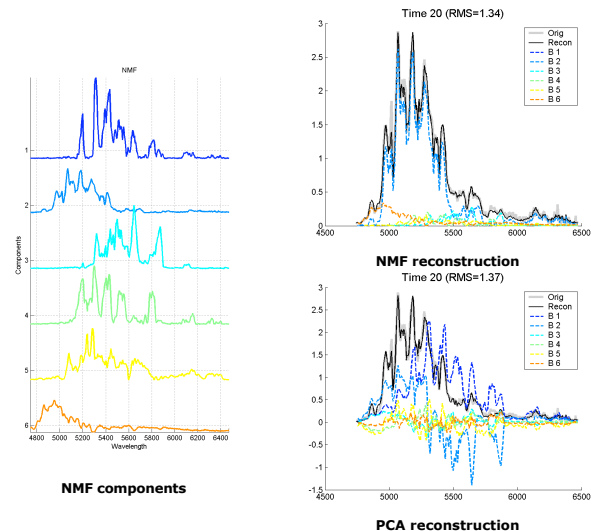
Decomposition of synthetic time series of SN Ia into 6 components via non-negative matrix factorization. Reconstructions have only slightly larger error, but show the contributions of components correlated with the phase of the SN, and do not cancel each other out due to the non-negativity constraint.

- Spectral features of the basis components reperesent physically meaningful phenomena, e.g. known absportion and emission lines.
- Proximity of spectra in the low-dimensional subspace matches known information about SNe Ia similarities, e.g. peculiar SNe are far from standard SNe, and synthetic SNe are well-matched to observed SNe.
- Classification of SNe using projections onto the basis result in clusters exhibiting high magnitude-redshift correlations.

More reconstructions.