Real-time Classification for The Palomar Transient Factory

Joseph Richards

UC Berkeley

Department of Astronomy
Department of Statistics

joeyrichar@gmail.com

Mini-Workshop on Computational AstroStatistics

Transient Classification Project (TCP)

Berkeley Astronomy:
Dan Starr, Dovi Poznanski, Maxime Rischard, Nat Butler, Chris Klein, Rachel Kennedy, Justin Huggins, Adam Morgan, Adam Miller, Josh Bloom

San Francisco State University:
John M. Brewer

Berkeley Statistics:
Sahand Negahban, James Long, Noureddine El Karoui, John Rice, Tamara Broderick, ...

Berkeley EECS:
Martin Wainwright, Massoud Nikravesh

Lawrence Berkeley Lab:
Peter Nugent, Horst Simon

Los Alamos Nat. Lab. / UC Santa Cruz:
Damian Eads

NSF grant 0941742

Cyber-enabled Discovery and Innovation (CDI)

Outline

1. Palomar Transient Factory (PTF)
2. Transient Classification Project (TCP)
   - DotAstro.org
   - Features for Classification
3. Light Curve Classification using Diffusion Map
   - Supernova Classification Challenge
   - Diffusion Map Classifier for PTF?

Palomar Transient Factory (PTF)

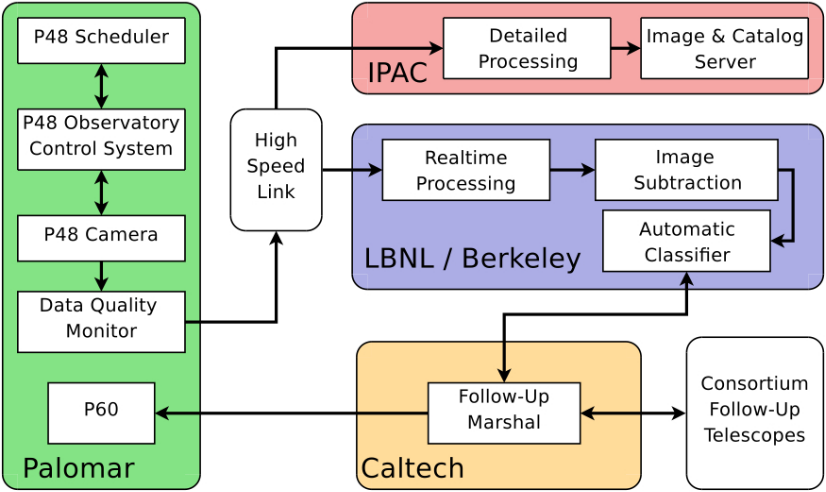
- Fully-automated survey to explore the optical transient sky
- Palomar 48" telescope (P48): 8.1 square-degree FoV, 101 megapixel CCD array, 2 filters

Experiment	**Science Goals**	**Setup**
5-day Cadence	SNe Ia, CC SNe, AGNs, QSOs, Novae, etc.	60 sec. exposures 8000 deg^2 / year
Dynamic Cadence	Unexplored (fast) regimes RR Lyr, CVs, flare stars, SNe, etc.	60 sec. exposures 1 min.- 5 day cadences
Orion Field	transiting planets	single field
Hα	Hα sky survey	narrow-band on/off Hα

- Follow-up of detected transients by Palomar 60" (P60) and telescopes at other facilities
- Follow-up decisions based on source classifications

Palomar Transient Factory (PTF)

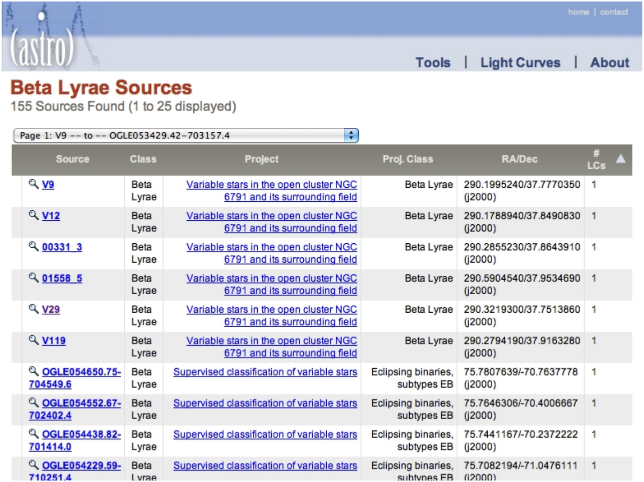
Law et al. (2009, PASP, 121, 1395)



Transient Classification Project (TCP)

- For optimal allocation of follow-up resources, need accurate, real-time classifications of transient events
- Some challenges:
  1. Need labeled data to train & validate our classifier

TCP: DotAstro.org



(astro)

home | contact

Tools | Light Curves | About

Beta Lyrae Sources

155 Sources Found (1 to 25 displayed)

Page 1: V9 -- to -- OGLE053429.42-703157.4

Source	Class	Project	Proj. Class	RA/Dec	# LCs
V9	Beta Lyrae	Variable stars in the open cluster NGC 6791 and its surrounding field	Beta Lyrae	290.1995240/37.7770350 (j2000)	1
V12	Beta Lyrae	Variable stars in the open cluster NGC 6791 and its surrounding field	Beta Lyrae	290.1788940/37.8490830 (j2000)	1
00331_3	Beta Lyrae	Variable stars in the open cluster NGC 6791 and its surrounding field	Beta Lyrae	290.2855230/37.8643910 (j2000)	1
01558_5	Beta Lyrae	Variable stars in the open cluster NGC 6791 and its surrounding field	Beta Lyrae	290.5904540/37.9534690 (j2000)	1
V29	Beta Lyrae	Variable stars in the open cluster NGC 6791 and its surrounding field	Beta Lyrae	290.3219300/37.7513860 (j2000)	1
V119	Beta Lyrae	Variable stars in the open cluster NGC 6791 and its surrounding field	Beta Lyrae	290.2794190/37.9163280 (j2000)	1
OGLE054650.75-704549.6	Beta Lyrae	Supervised classification of variable stars	Eclipsing binaries, subtypes EB	75.7807639/-70.7637778 (j2000)	1
OGLE054552.67-702402.4	Beta Lyrae	Supervised classification of variable stars	Eclipsing binaries, subtypes EB	75.7646306/-70.4006667 (j2000)	1
OGLE054438.82-701414.0	Beta Lyrae	Supervised classification of variable stars	Eclipsing binaries, subtypes EB	75.7441167/-70.2372222 (j2000)	1
OGLE054229.59-710251.4	Beta Lyrae	Supervised classification of variable stars	Eclipsing binaries, subtypes EB	75.7082194/-71.0476111 (j2000)	1

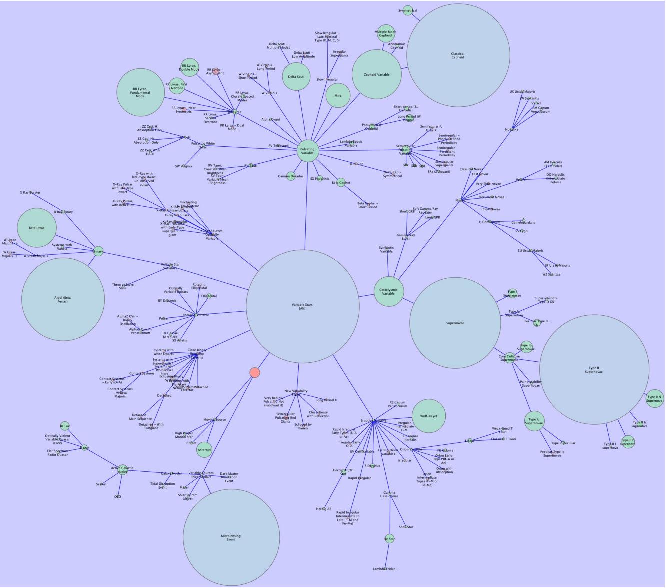
www.DotAstro.org: 100,000 sources, 150 classes

Astronomer-classified objects (from over 100 papers)

Transient Classification Project (TCP)

- For optimal allocation of follow-up resources, need accurate, real-time classifications of transient events
- Some challenges:
  1. Need labeled data to train & validate our classifier
  2. Highly multi-class problem with nested classifications

TCP: DotAstro.org Classification Taxonomy



Transient Classification Project (TCP)

- For optimal allocation of follow-up resources, need accurate, real-time classifications of transient events
- Some challenges:
  1. Need labeled data to train & validate our classifier
  2. Highly multi-class problem with nested classifications
  3. Light curves are noisy, irregularly sampled, and may include non-detections

Transient Classification Project (TCP)

- For optimal allocation of follow-up resources, need accurate, real-time classifications of transient events
- Some challenges:
  1. Need labeled data to train & validate our classifier
  2. Highly multi-class problem with nested classifications
  3. Light curves are noisy, irregularly sampled, and may include non-detections

Question: What information should we extract from the light curves to train our classifier?

TCP: Features for Classification

TCP currently uses:

1. Features derived from time-series light curves
   - flux quantiles, skewness, slope
   - period, amplitude of largest peaks of Lomb-Scargle periodogram
   - light-curve fitting parameters
   - color information
2. "Context" information
   - galactic latitude & longitude
   - distance from ecliptic plane
   - distance to nearest galaxy, properties of that galaxy

TCP: Features for Classification

TCP currently uses:

1. Features derived from time-series light curves
   - flux quantiles, skewness, slope
   - period, amplitude of largest peaks of Lomb-Scargle periodogram
   - light-curve fitting parameters
   - color information
2. "Context" information
   - galactic latitude & longitude
   - distance from ecliptic plane
   - distance to nearest galaxy, properties of that galaxy

Many possible features, lots of missing data, high levels of noise on some features!

Light Curve Classification using Diffusion Map

Light Curve Classification using Diffusion Map

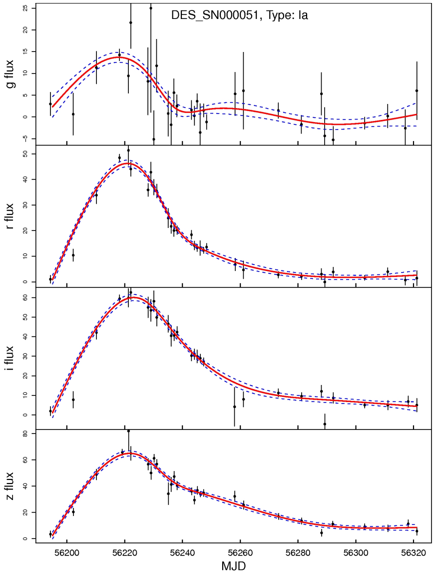
- Idea: First map each light curve, $\mathbf{x}$, into m-dimensional diffusion space $\mathbf{x} \mapsto \{\psi_1(\mathbf{x}), ..., \psi_m(\mathbf{x})\}$
- Use the diffusion map coordinates as features for a classifier
- Advantages:
  - All we need is distance between each pair of light curves, $s(\mathbf{x}_i, \mathbf{x}_j)$
  - Exploits the underlying sparse structure of the data set
  - Can avoid estimating physical parameters of each light curve

Supernova Classification Challenge

- DES SN Photometric Classification Challenge to test supernova light curve classification methods (Kessler et al. 2010, arXiv:1008.1024)
- *griz* SN light curves simulated from templates
- 1300 labeled SNe (Ia / II / Ibc) to classify 18,000

- Our entry: Find sparse structure in the SN database, exploit low-dimensional representation to build classifier
- InCA Group team: JWR, D. Homrighausen, C. Schafer, and P. Freeman

SN Photometric Classification: Methods



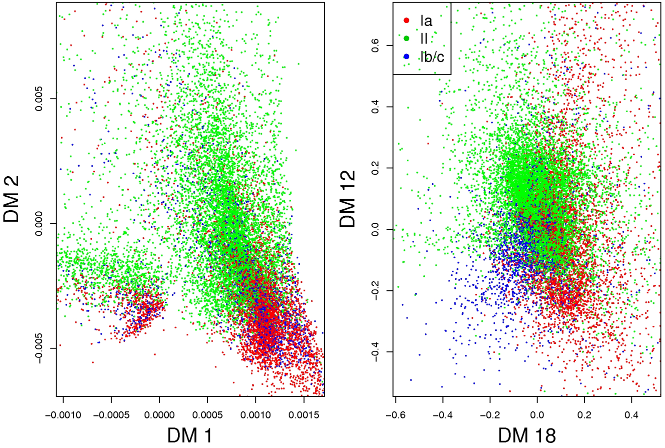
Data: Noisy realizations from SN LC templates in each of 4 filters

Our Approach:

1. Fit regression spline in each filter of each SN
2. Distances between SNe: weighted-ℓ_2 distance between normalized spline fits
3. Apply standard classification method using diffusion coords.

SN Photometric Classification: Results

Two-dimensional diffusion map representations of 18,000 SNe



Richards et al. (2010), in preparation

SN Photometric Classification: Results

SN Photometric Classification Challenge judged on both efficiency and purity of type Ia classifications.

$$\mathrm{FoM} = \frac{\mathrm{N}_{\mathrm{Ia}}^{\mathrm{true}}}{\mathrm{N}_{\mathrm{Ia}}^{\mathrm{Total}}} \times \frac{\mathrm{N}_{\mathrm{Ia}}^{\mathrm{true}}}{\mathrm{N}_{\mathrm{Ia}}^{\mathrm{true}} + 3\mathrm{N}_{\mathrm{Ia}}^{\mathrm{false}}}$$

SN Photometric Classification: Results

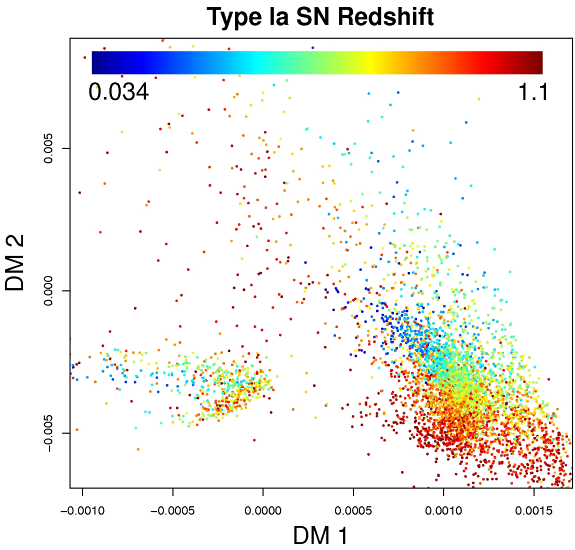
SN Photometric Classification Challenge judged on both efficiency and purity of type Ia classifications.

$$\mathrm{FoM} = \frac{N_{Ia}^{true}}{N_{Ia}^{Total}} \times \frac{N_{Ia}^{true}}{N_{Ia}^{true} + 3N_{Ia}^{false}}$$

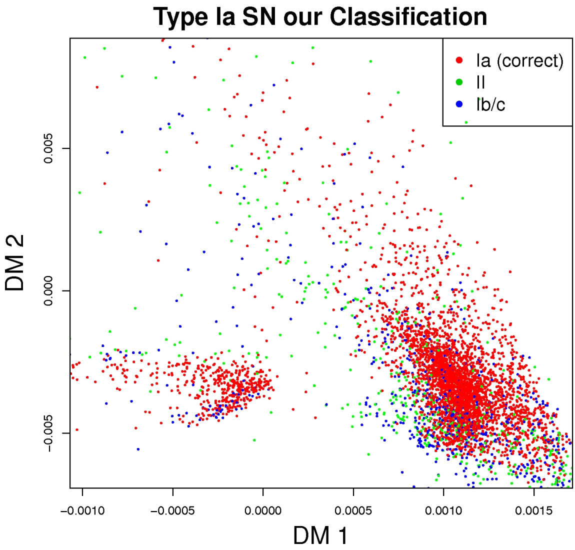
Submission	Description	Training FoM	Test FoM
this work	**Diffusion Map with Random Forest Classifier**	**0.810**	**0.236**
Rodney	Template Fitting	0.488	0.319
Gonzalez	Template Fitting	0.528	0.169
SNANA cuts	Template Fitting	0.582	0.221
Sako	Template Fitting	0.802	0.506
MGU+DU-1	Difference Boosting Neural Network on Light Curve Slopes	0.385	0.079
MGU+DU-2	Random Forest on Light Curve Slopes	0.635	0.148
JEDI KDE	Kernel Density Estimation with 21 Parameters*	0.946	0.371
JEDI Boost	Boosted Decision Trees with 21 Parameters*	0.961	0.204
All Ia	Classify all SNe as Type Ia	0.385	0.107

*light curve for each filter is fit to a modified Γ-distribution function with five parameters

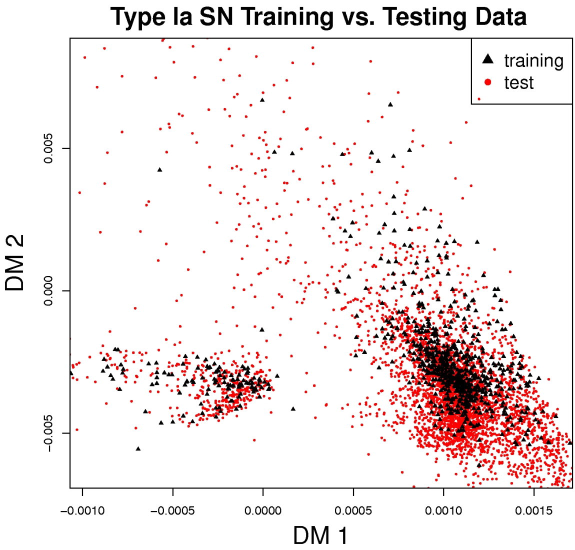
SN Photometric Classification: Redshift



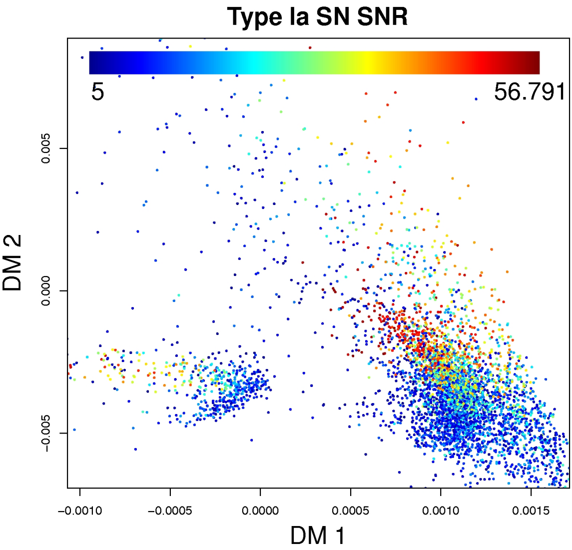
SN Photometric Classification: Predicted Class



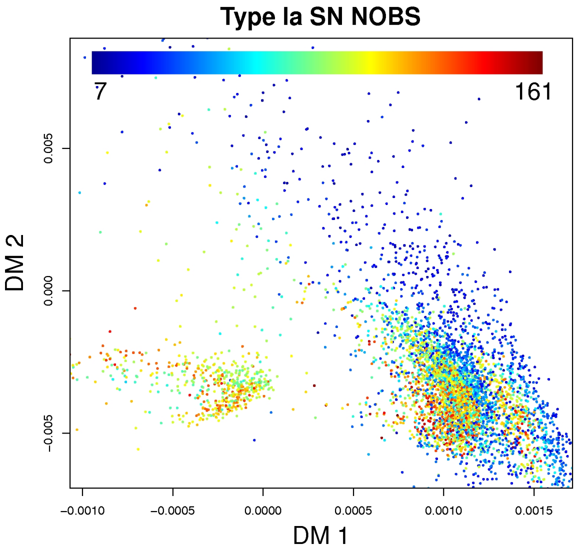
SN Photometric Classification: Training vs. Test



SN Photometric Classification: SNR



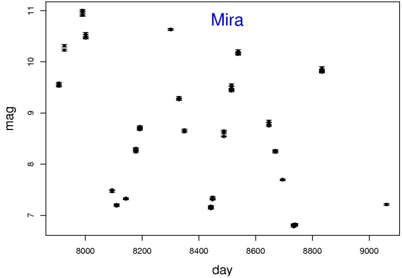
SN Photometric Classification: No. of Obs.

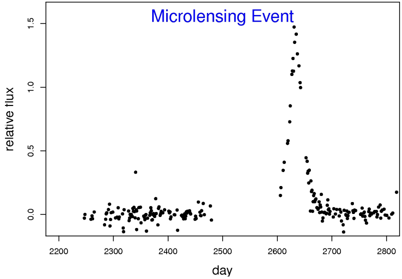


SN Photometric Classification: Lessons Learned

- Diffusion map is a competitive method for SN classification
- Latent variables (e.g. redshift) are captured by the diffusion coordinates.
  - No time dilation or K-corrections!
  - Can be used as photo-z estimator
- Need more SN training data at high-z!
  - If spectral coverage is not possible, can we model the physics?
- Much care is needed in choosing $s(\cdot, \cdot)$

Diffusion Map for PTF Classifier: Challenges





- Must deal with periodic (e.g. variable stars) and non-periodic (e.g. explosive events) light curves
- How do we construct a **general distance measure**, $s(\cdot,\cdot)$?
- Flux normalization? Time zero-point? Period folding? Fourier space?

Summary

- Accurate, real-time transient classification is imperative for the PTF (and LSST)
- `DotAstro.org` contains expert-classified light curves for 100,000 sources
- Classification problem is highly multi-class and data are noisy and heterogeneous
- What are the optimal features for classification? Can we incorporate measurement errors?
- Diffusion map can be used to uncover simple structure in sets of light curves
    - Application to SN Classification Challenge yielded impressive results
- Can we adapt this methodology to build a general transient classifier?

TCP Publications

Bloom, J. S., Starr, D. L., Butler, N. R., Nugent, P., Rischard, M., Eads, D., Poznanski, D. **Towards a real-time transient classification engine** (2008, AN, 329, 284)

Starr, Dan L., Bloom, J. S., Butler, N. R. **Real-time Transient Classification Pipeline** (2008, AIPC, 1000, 635)

Bloom, Joshua S., Starr, D. L., Butler, N. R., Poznanski, D., Rischard, M., Kennedy, R., Brewer, J. **Rapid and Automated Classification of Events from the Palomar Transient Factory** (2009, AAS, 41, 419)

Poznanski, Dovi et al. **The Standard Candle Method For Type II-P Supernovae - New Sample And PTF Perspective** (2009, AAS, 41, 419)

Brewer, J. M., Bloom, J. S., Kennedy, R., Starr, D. L. **A Web-Based Framework For a Time-Domain Warehouse** (2009, ASPC, 411, 357)

Starr, D. L., Bloom, J. S., Brewer, J. M., Butler, N. R., Poznanski, D., Rischard, M., Klein, C. **The Berkeley Transient Classification Pipeline: Deriving Real-time Knowledge from Time-domain Surveys** (2009, ASPC, 411, 493)

Butler, Nathaniel R., Bloom, Joshua S. **Optimal Time-Series Selection of Quasars** (2010, arXiv:1008.3143)