Automated classification of transients

Ashish Mahabal
aam@astro.caltech.edu
7 June 2011
KISS, Digging Deeper workshop tutorial
Collaborators

• Caltech
  – George Djorgovski
  – Ciro Donalek
  – Andrew Drake
  – Matthew Graham
  – Roy Williams
  – Nihar Sharma
  – Yutong Chen

• JPL
  – Baback Moghaddam
  – Mike Turmon

Plus at various other institutes all over, but especially in US, India and Italy
What is a transient?

Example: 4 individual exposures, separated by 10 min

Fast transient (flaring dM), CSS080118:112149–131310

Light curve

Something that has a large delta-magnitude for a small delta-time
Examples of CRTS Transients

CSS090429:135125-075714
Flare star

CSS090429:101546+033311
Dwarf Nova

CSS090426:074240+544425
Blazar, 2EG J0744+5438

Vastly different physical phenomena, and yet they look the same!
Which ones are the most interesting and worthy of follow-up?

Rapid, automated transient classification is a critical need!
SED: Spectral Energy Distribution

http://hera.ph1.uni-koeln.de/~heintzma/k1/3C273_HST.htm

Ashish Mahabal
Colors (e.g. from follow-up from Palomar 60-inch) are a narrower aspect of the SED.
Semantic Tree of Astronomical Variables and Transients

The diagram outlines the classification of astronomical variables and transients into several categories:

- **Asteroids**
  - Rotation
  - Eclipse

- **Stars**
  - Eclipse
  - Microlensing
  - Rotation

- **Extrinsic**
  - ELL
  - BY Dra
  - UV Ceti

- **Intrinsic**
  - Red dwarfs (K-M stars)

- **SN Subtypes**
  - Mostly optical viewpoint

- **AGN Subtypes**
  - Mostly optical viewpoint

- **SN Subtypes**
  - Mostly optical viewpoint

- **Cataclysmic**
  - ZAND
  - Symbiotic

- **Eruptive**
  - SN Supernovae

- **Pulsation**
  - Dwarf novae

- **Secular**
  - Solar-like

- **Planetary transits**

Credit: L. Eyre & N. Mowlavi (10/2007)
The Palomar-Quest (PQ) Digital Synoptic Sky Survey

- Palomar 48-in. + 112-CCD, 161 Mpix camera
- A Caltech-Yale collab. Co-PIs: C. Baltay & SGD; plus other groups worldwide (LBL, etc.)
- Many passes with up to 4 filters (UBRI/griz), time baselines from minutes to years
- Collected > 50 TB of data
- **Key goal:** Exploration of the time domain

http://palquest.org

↔ LBL SNF search (Nugent et al.)

> 700 SNe discovered
Catalina Sky Survey(s):

CRTS uses the data from all three Catalina NEO surveys, with a coverage of up to 2,500 deg$^2$ / night, and the total area coverage of $\sim 30,000$ deg$^2$

---

We are processing the Catalina data streams in real time to look for astrophysical transients
Recent, current and future multiepoch surveys

Hundreds of thousands of transients per night in the near future

Orders of magnitudes different.

Move towards digital movies!

Ashish Mahabal
CRTS Event Detections

Distinct Events Detection Statistics as of 5 Jun 2011 UT:

<table>
<thead>
<tr>
<th>Tel</th>
<th>All OTs</th>
<th>SNe</th>
<th>CVs</th>
<th>Blazars</th>
<th>Ast/flares</th>
<th>CV/SN</th>
<th>AGN</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSS</td>
<td>2033</td>
<td>596</td>
<td>501</td>
<td>113</td>
<td>184</td>
<td>275</td>
<td>229</td>
<td>195</td>
</tr>
<tr>
<td>MLS</td>
<td>1560</td>
<td>183</td>
<td>38</td>
<td>12</td>
<td>122</td>
<td>374</td>
<td>744</td>
<td>214</td>
</tr>
<tr>
<td>SSS</td>
<td>227</td>
<td>24</td>
<td>93</td>
<td>7</td>
<td>5</td>
<td>43</td>
<td>16</td>
<td>42</td>
</tr>
<tr>
<td>Total</td>
<td>3820</td>
<td>803</td>
<td>632</td>
<td>132</td>
<td>311</td>
<td>692</td>
<td>989</td>
<td>451</td>
</tr>
</tbody>
</table>

- Threshold set deliberately very high – only the most dramatic transients are pulled out in the real time
- About 1 strong transient per $10^6$ source detections
- The rate of significant transients/variables is at least an order of magnitude higher
- Many events are re-detected repeatedly (not counted above)
The Palomar-Quest Event Factory

Detect $\sim 1 - 2 \times 10^6$ sources per half-night scan

Compare with the baseline sky

Find $\sim 10^3$ apparent transients (in the data)

Remove instrum. artifacts

Identify $\sim 2 - 4 \times 10^2$ real transients (on the sky)

Remove asteroids

Identify $\sim 1 - 10$ possible Astrophysical transients

Comparison with the baseline sky

Classification and follow-up

Ashish Mahabal
Towards Automated Event Classification

A necessity for large synoptic surveys

Event parameters: \( m_1(t), m_2(t), \ldots \), \( \alpha, \delta, \mu, \ldots \), image shape...

Event Classification Engine

Priors:
- \( P(SN\ Ia) = \ldots \)
- \( P(SN\ II) = \ldots \)
- \( P(AGN) = \ldots \)
- \( P(CV) = \ldots \)
- \( P(dM) = \ldots \)

Classification probabilities (evolving, iterated)

With M Turmon and B Moghaddam, JPL

Expert and ML generated priors

- colors
- lightcurves

6/7/11
Sample Light Curves

Blazar PKS0823+033  CV 111545+425822

Variables and transients – the distinction is one of perception, and your aims
Building Bayesian Networks

• Handling of incomplete data
  – Real-world cases

• Learning causal connections
  – What variable caused what

• Incorporating domain knowledge
  – Experts can weight in at different points

• Memorizing (aka overfitting) avoided
  – No holdout necessary
BN is an Acyclic Directed Graph

Once certain characteristics are observed some sections become irrelevant
Sample data input to BN

The output is BN class which is fed to skyalert as an annotation to the original event

CV/SN classification ~80% with single epoch
Naïve Bayes

$$P(y = k \mid x) = P(x \mid y = k)P(k) / P(x) \propto P(k)P(x \mid y = k) \approx P(k) \prod_{b=1}^{B} P(x_b \mid y = k)$$

- $x$: feature vector of event parameters
- $y$: object class that gives rise to $x$ (1<y<k)
- Certain features of $x$ known: (position, flux)
- Others will be unknown: (color, delta-mag)
- Assumption: based on $y$, $x$ is decomposable into $B$ distinct independent classes (labeled $x_b$)
- This helps with the curse of dimensionality
- Also allows us to deal with missing values
The importance of context

Which galaxy does a supernova belong to?

The need to see the big picture

Matthew Graham: Semantics
Characterization Vs. Classification

• Early focus on the extraction and dissemination of time series

• Characterizations is important
  – \( \frac{dm}{dt} \)
  – change of direction per unit time
  – change in periodicities (e.g., wavelet or fourier decomposition);
  – variation in \( \frac{dm}{dt} \)
  – acceleration in \( \frac{dm}{dt} \)

Most SNe will not become fainter and then brighten up
Aspects of “Gap” processing

• Gap features capture sparse or irregular LCs
• The features, and thus the underlying density models, are invariant to absolute magnitude and time shifts
• Features & densities allow bound-only flux observations
  – Under poor seeing, we obtain only bounds like $m > 18$
Decision Tree decomposes this multi-class classifier into a series of binary discrimination tasks.

This specific DT follows the stratification that seems natural to astronomers.

All nodes shown were implemented via “gap histogram” binary classifiers.
By taking subsections of $dt/dm$ space determine which area is characteristic for which kind of variable
For a specific dm/dt region

```bash
>> prep_Blazar.CV
>> RUNME_prototype_classifier
```

- Blazar accuracy = 83.33 %
- CV accuracy = 58.85 %
- Total accuracy = 64.31 %
- Average accuracy = 71.09 %

6/7/11
Ashish Mahabal
Follow-Up Observations:

- **Photometry** (P60, NMSU, DAO, HTN, India, Mexico, etc.)
- **Spectroscopy** (Gemini N+S, Keck, P200, SMARTS, IGO, MDM)

CSS090421:174806+340401  A blazar, also monitored at OVRO in radio
Automating the Optimal Follow-Up

What type of follow-up data has the greatest potential to discriminate among the competing models (event classes)?

Request follow-up observations from the optimal available facility

Collaboration with B. Moghaddam, M. Turmon (JPL)
Bayesian Network/fusion modules are no Cartesian theatre

• Different parameters, methods are separate (though perhaps not independent) probes

(non-)Cartesian theatre
One observation can drive the direction given the large number of possible candidates
Not much scope for error
Event Publishing / Dissemination
skyalert.org

PI: R. Williams

- Real time:
  - VOEvents, Twitter, iApp (thousands of events)
  - Also on SkyAlert.org, feeds to the WWT, GoogleSky

- Next day: annotated tables on the CRTS website

<table>
<thead>
<tr>
<th>CSS ID</th>
<th>RA (J2000)</th>
<th>Dec (J2000)</th>
<th>Date</th>
<th>Mag</th>
<th>CSS images</th>
<th>SDSS</th>
<th>Others</th>
<th>Followed</th>
<th>Last</th>
<th>LC</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSS091121:221159+263906</td>
<td>332.99697</td>
<td>26.65153</td>
<td>20091121</td>
<td>18.33</td>
<td>911211261084134848</td>
<td>no</td>
<td>34848</td>
<td>no</td>
<td>2009-11-21</td>
<td>34848</td>
<td>SN/Blazar mag 21</td>
</tr>
<tr>
<td>CSS091121:013728+253450</td>
<td>24.36768</td>
<td>25.58061</td>
<td>20091121</td>
<td>17.78</td>
<td>911211260084103595</td>
<td>no</td>
<td>03595</td>
<td>no</td>
<td>2009-11-21</td>
<td>03595</td>
<td>SN/CV</td>
</tr>
<tr>
<td>CSS091121:032627+070744</td>
<td>51.61384</td>
<td>7.12902</td>
<td>20091121</td>
<td>16.66</td>
<td>911211607201914436</td>
<td>no</td>
<td>24436</td>
<td>no</td>
<td>2009-11-21</td>
<td>24436</td>
<td>CV mag 21</td>
</tr>
<tr>
<td>CSS091121:033232+020439</td>
<td>53.13295</td>
<td>2.07747</td>
<td>20091121</td>
<td>16.93</td>
<td>911211010109144343</td>
<td>no</td>
<td>34434</td>
<td>no</td>
<td>2009-11-21</td>
<td>34434</td>
<td>CV mag 20</td>
</tr>
<tr>
<td>CSS091121:085600-051945</td>
<td>133.99922</td>
<td>-5.32906</td>
<td>20091121</td>
<td>18.17</td>
<td>911210040484107252</td>
<td>no</td>
<td>07252</td>
<td>no</td>
<td>2009-11-21</td>
<td>07252</td>
<td>SN CFHT mag 22 gal</td>
</tr>
<tr>
<td>CSS091120:100525+511639</td>
<td>151.35223</td>
<td>51.27742</td>
<td>20091120</td>
<td>18.80</td>
<td>911201520354008835</td>
<td>yes</td>
<td>08835</td>
<td>no</td>
<td>2009-11-20</td>
<td>08835</td>
<td>SN SDSS mag 21.9 gal</td>
</tr>
<tr>
<td>CSS091120:082908+482639</td>
<td>127.28503</td>
<td>48.44423</td>
<td>20091120</td>
<td>15.69</td>
<td>911201490314109371</td>
<td>yes</td>
<td>09371</td>
<td>no</td>
<td>2009-11-20</td>
<td>09371</td>
<td>CV/SN SDSS mag 21.6 gal?</td>
</tr>
<tr>
<td>CSS091120:004417+411864</td>
<td>11.07004</td>
<td>41.31494</td>
<td>20091120</td>
<td>17.00</td>
<td>91120140004145995</td>
<td>yes</td>
<td>45995</td>
<td>no</td>
<td>2009-11-20</td>
<td>45995</td>
<td>Nova M31 2009-11d</td>
</tr>
<tr>
<td>CSS091120:001019+410455</td>
<td>2.58044</td>
<td>41.08191</td>
<td>20091120</td>
<td>16.69</td>
<td>91120140001437919</td>
<td>no</td>
<td>37919</td>
<td>no</td>
<td>2009-11-20</td>
<td>37919</td>
<td>CV mag 20.0</td>
</tr>
</tbody>
</table>
Real Time Event Publishing via VOEvents and SkyAlert

From the CRTS stream. Catalina Real-time Transient Survey. Position is $115.98635, 21.1753 \pm 0.0012$. This portfolio initiated 2009-11-11 08:35:18.

Subscribe to VOEvents via email, RSS, Atom feed, etc.

Basic event info

Linked VO/archival data for classif. and follow-up

Dynamically growing portfolio

6/7/11 Ashish Mahabal
http://www.skyalert.org/

- Subscribe to feeds – set-up alerts based on your own criteria
- Suppose you subscribe to a “CRTS + P60” alert
  - get the parameters that you want
  - be able to ‘resolve’ the event into its portfolio
- wget –O myevent
  http://www.skyalert.org/params/ivo:+nvo.caltech+voeventnet+catot;902200180274145987/
Reading json output from skyalert

x <- fromJSON(file = "myevent")
X

That gives:...
$portfolio$`ivo://nvo.caltech/voeventnet/csFollowup#observation`$nearest_obj_usnob
$portfolio$`ivo://nvo.caltech/voeventnet/csFollowup#observation`$nearest_obj_usnob$distance
[1] "0.5"

x$portfolio$`ivo://nvo.caltech/voeventnet/sdssFollowup#observation`[[1]]$ISOtime
gives
Given a reported coincidence:
-- Gravitational wave trigger with time and probability density
-- Burst detection with time and position

Questions:
-- What chance these are physically connected?
-- Should we immediately slew telescopes?
DAME is a joint effort between University Federico II, INAF-OACN, and Caltech aimed at implementing (as web application) a scientific gateway for data analysis, exploration, mining and visualization tools, on top of virtualized distributed computing environment.

http://voneural.na.infn.it/
Technical and management info
Documents
Science cases
Newsletter

http://143.225.93.239:8080/MyDameFE/
Web application PROTOTYPE
(ALPHA release)

Ashish Mahabal

brescia@na.astro.it
K-MEANS PARTITIONING
Click here to view Plot

sizes:
33 30 73 42 46 42

centers:

<table>
<thead>
<tr>
<th>Xpos</th>
<th>Ypos</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1186.9091</td>
<td>1846.9091</td>
</tr>
<tr>
<td>2 514.2000</td>
<td>1866.8000</td>
</tr>
<tr>
<td>3 1578.8219</td>
<td>730.3562</td>
</tr>
<tr>
<td>4 568.3333</td>
<td>565.4286</td>
</tr>
<tr>
<td>5 1760.1304</td>
<td>1554.7174</td>
</tr>
<tr>
<td>6 542.5000</td>
<td>1301.9048</td>
</tr>
</tbody>
</table>

withinss:
2292505 1784038 10275335 35:
Harvesting the Human Pattern Recognition

Recognizing the artifacts (false transients)

*Contextual information is essential*

A more sophisticated case uses a **prior (expert) knowledge**: Star-like transient apparently associated with a non-coincident galaxy a likely Supernova

Spiral host galaxy a possible Type II

*How to capture this and teach a machine to do the same thing?*
Use Case Scenario:
Light curves are ambiguous, but the presence of a possible host galaxy suggest that it is a Supernova.

6/7/11
Ashish Mahabal
AstroCollation: Towards Harvesting Human Pattern Recognition and Domain Expertise

Input Event Streams
- CRTS
- SkyAlert

User interface
- Visualization
- UI database

Process evaluation and analysis

Event manager

External Data Archives, VO

Event interface

Algorithms
- Semantics
- Synthesis engine

Machine learning

Lead:
M. Graham
Transient classification mantra

- Obtain a couple of epochs in one or more filters
- Assigns probabilities for different classes
- Choose observations (filters, wavelengths) for best discrimination
- Feed the new observations back in
- Revise probabilities, choose observations, ... 
- Based on confirmed class revise priors

Bayesian network, Gap processing, (DAME, VOStat, VO), Skyalert
Video pictures ala sixth sense: Portfolios of transients (or any object for that matter) – automagically updating lightcurves, SEDs etc.

The Glass Bead Game of astronomy – connecting everything possible for classifying transients in realtime

Summary

Future: Minority Report like interfaces in open sims.