Combining Diverse Classifiers [for astronomical transients] {and in real-time}

Ashish Mahabal, Caltech
IJCAI, 16 July 2009
SG Djorgovski, Ciro Donalek, Baback Moghaddam, Michael Turmon, Andrew Drake, Matthew Graham, Roy Williams, ...
• Colors and context info with BNs
• Lightcurves and GPRs
• Combining them
Transformations in astronomy

- Shallow to deep
- Small to large
- Sporadic to repeated
Towards digital movies ...

ROTSE, NEAT, DLS, FSVS, ...
DPOSS, PQ, CSS, PTF, Pan-STARRS, LSST ...
(GALEX, Spitzer, FIRST, ..., SKA)

• Area covered
• Depth of coverage
• Number of wavelengths
• Baseline in time
• Number of epochs

*Etendue* (throughput) $\sim dA \ast dm \ast d\lambda \ast dt \ast dn$
Synoptic skysurveys

Opened up new dimensions

Challenges besides data mining:
• Lots of follow-up observing
• Selecting candidates to follow
• In real-time
Bayesian Networks to the rescue

• BNs of various flavors can tackle these issues (but there are many barriers to be crossed!)

• Data define network
• No “training” necessary
Colors for classification

- Magnitude as basic observation (flux)
- Color as flux ratio
- Color-color diagram as a diagnostic
P60 follow-up of a CRTS transient

http://www.astro.caltech.edu/P60FollowUp/
Class can be a function of time
DAG for variables like QMR-DT

Variables and observed properties
CV, SN, Blazars, Rest

spectra

Incidental parameters

-4 -> 4 (10 bins each)

Phenomenology

Class

Colors

Other observed parameters

lightcurves
8% CV classified as SN, 60% of objects classified as CV are actually CV

<table>
<thead>
<tr>
<th>3 colors, no gb (WTA)</th>
<th>CV (0.60)</th>
<th>SN (0.73)</th>
<th>BL (0.30)</th>
<th>REST (0.22)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV</td>
<td>0.72</td>
<td>0.08</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td>SN</td>
<td>0.42</td>
<td>0.41</td>
<td>0.08</td>
<td>0.09</td>
</tr>
<tr>
<td>BL</td>
<td>0.30</td>
<td>0.00</td>
<td>0.49</td>
<td>0.21</td>
</tr>
<tr>
<td>REST</td>
<td>0.32</td>
<td>0.13</td>
<td>0.34</td>
<td>0.21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3 colors + gb (WTA)</th>
<th>CV (0.65)</th>
<th>SN (0.71)</th>
<th>BL (0.33)</th>
<th>REST (0.23)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV</td>
<td>0.72</td>
<td>0.08</td>
<td>0.08</td>
<td>0.13</td>
</tr>
<tr>
<td>SN</td>
<td>0.23</td>
<td>0.46</td>
<td>0.12</td>
<td>0.19</td>
</tr>
<tr>
<td>BL</td>
<td>0.24</td>
<td>0.03</td>
<td>0.49</td>
<td>0.24</td>
</tr>
<tr>
<td>REST</td>
<td>0.34</td>
<td>0.18</td>
<td>0.21</td>
<td>0.26</td>
</tr>
<tr>
<td>Type</td>
<td>CV</td>
<td>SN</td>
<td>BL</td>
<td>REST</td>
</tr>
<tr>
<td>----------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>------</td>
</tr>
<tr>
<td>3,n,w</td>
<td>0.60</td>
<td>0.73</td>
<td>0.30</td>
<td>0.22</td>
</tr>
<tr>
<td>3,y,w</td>
<td>0.65</td>
<td>0.71</td>
<td>0.33</td>
<td>0.23</td>
</tr>
<tr>
<td>2+,n,w</td>
<td>0.58</td>
<td>0.77</td>
<td>0.31</td>
<td>0.17</td>
</tr>
<tr>
<td>2+y,w</td>
<td>0.65</td>
<td>0.78</td>
<td>0.41</td>
<td>0.19</td>
</tr>
<tr>
<td>3,y,50</td>
<td>0.75</td>
<td>0.83</td>
<td>0.37</td>
<td>0.19</td>
</tr>
<tr>
<td>3,y,40-10</td>
<td>0.73</td>
<td>0.82</td>
<td>0.37</td>
<td>0.18</td>
</tr>
<tr>
<td>2+,y,50</td>
<td>0.74</td>
<td>0.88</td>
<td>0.43</td>
<td>0.19</td>
</tr>
</tbody>
</table>
• More context information will help
• Heterogeneity
• How to maintain uniformity [from the same source]?
• What when filters change?
• Uniformity of priors?
  • Number of objects
  • Their magnitude range
  • Spread over time
• Ground truth?
Questions raised by the Data paucity regime

• How many classes?
• Too few: probabilities incorrect (where do objects belonging to unrepresented classes go?)
• Too many: overlaps increase (e.g. SN of different types; variables of different types) and probability splits into smaller fractions
• What kind of winner?
• May be evaluate probability for each class independently?
• With GPR something like that may be possible.
Given several epochs and corresponding magnitudes, estimate the likelihood of a particular magnitude for a new epoch (using some covariance function)
Graph of a SN lightcurve ready to be fitted using GPR (using Squared exponential covariance function from Matlab’s GPML).

\[ Cov(f(x_p), f(x_q)) = k_y(x_p, x_q) = \sigma_f^2 e^{-\frac{1}{2} l^2 (x_p - x_q)^2} + \sigma_n^2 \delta_{pq} \]

The 3 hyperparameters are “free” and are varied.
Graph of a SN lightcurve fitted using GPR (using Squared exponential covariance function from Matlab’s GPML).

$$Cov(f(x_p), f(x_q)) = k_y(x_p, x_q) = \sigma_f^2 e^{-\frac{1}{2}l^2(x_p-x_q)^2} + \sigma_n^2 \delta_{pq}$$

The 3 hyperparameters are “free” and are varied.
Figure 3: Estimation of $y_*$ (solid line) for a function with (a) short-term and long-term dynamics, and (b) long-term dynamics and a periodic element. Observations are shown as crosses.

$$k(x, x') = \sigma_f^2 \exp \left[ \frac{-(x - x')^2}{2l_1^2} \right] + \sigma_f^2 \exp \left[ \frac{-(x - x')^2}{2l_2^2} \right] + \sigma_n^2 \delta(x, x')$$

$$k(x, x') = \sigma_f^2 \exp \left[ \frac{-(x - x')^2}{2l^2} \right] + \exp\{-2 \sin^2[\nu \pi (x - x')]\} + \sigma_n^2 \delta(x, x')$$
Graph of a Mira lightcurve fitted using GPR (using a function that has a periodic component).
Mira star classifier results, sample_size=4, samples_per_graph=10

Samples from file (every increment by 1 indicates a new file)
Question of normalization

• First epoch after detection (for SN, at what distance from peak?)
• Number and frequency of epochs
• Periodic but non-constant periods
Using dMags and dT

dmag for all points in Supernovae Type Ia lightcurves

Graph of expected data using fixed hyperparameters: (4.02, 1.0, 0.1)

dt for all points in Supernovae Type Ia lightcurves
• Combining lightcurves + colors + contextual info in the right proportion can be tricky (just like the WTA, 50%, 40-10% classifications)
Fusion Module

Next steps include combining CRTS, P60, PTF etc. data

Next steps include combining CRTS, P60, PTF etc. data
Follow-up (for missing values)

• Such that it will help discriminate better
• Serve probabilities so that consumers can choose their types of transients
• Widest possible models
• (but we need a proper listing of all follow-up resources)
• (and they need to be uniformly connected and talking to each other)
Choosing follow-up configs

r-i color, hi-z quasar, blue star
Summary questions

• Number of classes (may have to be hierarchical)
• Normalization
• Combining diverse probabilities

Preprocessing + processing = incorporating domain knowledge and knowledge about the nature of data right into the DM/AI methodologies
Classification based on minimal data -> killer app.

Synoptic, panoramic surveys → Event discovery
Rapid follow-up and multi-λ → Keys to understanding
Early classification → Selective follow-up

A very rich variety of astrophysical phenomena: from asteroids to cosmology, from extrasolar planets to extreme relativistic physics

All interesting things are outliers in some parameter space. Event discovery is just a start: 99% of the astrophysics is in the follow-up, and classification