Complete Classification Conundrum

[Graph showing data points and a trend line]

Ashish Mahabal
aam@astro.caltech.edu
Center for Data-Driven Discovery (CD^3), Caltech
Collaborators: CRTS, PTF, LSST, SAMS, IUCAA ... teams
SCMA VI, CMU, 20160607
Sky Maps of a few (optical) surveys

CRTS

PTF

Gaia

LSST
Stars, Milky Way, and Local Volume

Solar System

Statistics and Informatics

Galaxies

Dark energy

STRONG LENSING

Active Galactic Nuclei

Transients and Variable Stars

Large Scale Structure/Baryon Oscillation
Stars, Milky Way, and Local Volume

Solar System

Statistics and Informatics

Galaxies

Dark energy

Strong Lensing

Active Galactic Nuclei

Transients and Variable Stars

Large Scale Structure/Baryon Oscillation
What is a transient?

One that has a **large brightness change** (delta-magnitude) within a short timespan  (small delta-time)
Challenge 1: Characterize/Classify as much with as little data as possible

Despite the heterogeneity, gaps, heteroskedasticity
Challenge 1: Characterize/Classify as much with as little data as possible

Credit: L. Eyer & N. Mowlavi (03/2009)  
(updated 04/2013)
Example CRTS Transients

CSS090429:135125-075714  Flare star

CSS090429:101546+033311  Dwarf Nova

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

Different phenomena look the same!
AGN Variability - different perspectives

CRTS

Kepler
Truncation and Censoring
What You See Is All There Is (WYSIATI)

When regressing base rates should not be forgotten.

Faraway, Mahabal et al. 2015
500 Million Light Curves with \( \sim 10^{11} \) data points

RR Lyrae

W Uma

Eclipsing

Flare star (UV Ceti)

CV

Blazar

CRTS PIs Djorgovski, Drake
Challenge 2: Only a small fraction are rare - find/characterize them early

CRTS 10+ year status

<table>
<thead>
<tr>
<th>Telescope</th>
<th>All OTs</th>
<th>Supernovae</th>
<th>Cataclysmic Variables</th>
<th>Blazars</th>
<th>Asteroids/Flares</th>
<th>CV or SN</th>
<th>AGN</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSS</td>
<td>5353</td>
<td>1669</td>
<td>964</td>
<td>265</td>
<td>366</td>
<td>562</td>
<td>640</td>
<td>977</td>
</tr>
<tr>
<td>MLS</td>
<td>5879</td>
<td>886</td>
<td>119</td>
<td>109</td>
<td>299</td>
<td>890</td>
<td>2787</td>
<td>1004</td>
</tr>
<tr>
<td>SSS</td>
<td>700</td>
<td>105</td>
<td>256</td>
<td>18</td>
<td>13</td>
<td>109</td>
<td>33</td>
<td>171</td>
</tr>
<tr>
<td>SNhunt</td>
<td>197</td>
<td>197</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>12129</td>
<td>2857</td>
<td>1339</td>
<td>392</td>
<td>678</td>
<td>1561</td>
<td>3460</td>
<td>2152</td>
</tr>
</tbody>
</table>

Current Status: Few tens of transients per night
Future (LSST): $10^6 - 10^7$ per night; $10^4$ per minute
That is why we need automatic classification algorithms
Variability on huge range of timescales

<table>
<thead>
<tr>
<th>Class</th>
<th>Timescale</th>
<th>Amplitude (Δmags)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WD Pulsations</td>
<td>4-10 min</td>
<td>0.01 - 0.1</td>
</tr>
<tr>
<td>AM CVn (orbital period)</td>
<td>10-65 min</td>
<td>0.1 - 1</td>
</tr>
<tr>
<td>WD spin (int. polars)</td>
<td>20-60 min</td>
<td>0.02 - 0.4</td>
</tr>
<tr>
<td>AM CVn outbursts</td>
<td>1-5 days</td>
<td>2 - 5</td>
</tr>
<tr>
<td>Dwarf Novae outburst</td>
<td>4 days - 30 years</td>
<td>2 - 8</td>
</tr>
<tr>
<td>Symbiotic (outburst)</td>
<td>weeks-months</td>
<td>1 - 3</td>
</tr>
<tr>
<td>Novae-like high/low</td>
<td>days-years</td>
<td>2 - 5</td>
</tr>
<tr>
<td>Recurrent Novae</td>
<td>10-20 year</td>
<td>6 - 11</td>
</tr>
<tr>
<td>Novae</td>
<td>$10^3\text{-}10^4$ yr</td>
<td>7 - 15</td>
</tr>
</tbody>
</table>

Slide from Lucianne Walkowicz
## Expected Rate of Transients

<table>
<thead>
<tr>
<th>Class</th>
<th>Mag</th>
<th>t (days)</th>
<th>Universal Rate</th>
<th>LSST Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luminous SNe</td>
<td>-19...-23</td>
<td>50 - 400</td>
<td>$10^{-7}$ Mpc$^{-3}$ yr$^{-1}$</td>
<td>20000</td>
</tr>
<tr>
<td>Orphan Afterglows SHB</td>
<td>-14...-18</td>
<td>5 - 15</td>
<td>$3 \times 10^{-7}...-9$ Mpc$^{-3}$ yr$^{-1}$</td>
<td>~10 - 100</td>
</tr>
<tr>
<td>Orphan Afterglows LSB</td>
<td>-22...-26</td>
<td>2 - 15</td>
<td>$3 \times 10^{-10}...-11$ Mpc$^{-3}$ yr$^{-1}$</td>
<td>1000</td>
</tr>
<tr>
<td>On-axis GRB afterglows</td>
<td>-37</td>
<td>1 - 15</td>
<td>$10^{-11}$ Mpc$^{-3}$ yr$^{-1}$</td>
<td>~50</td>
</tr>
<tr>
<td>Tidal Disruption Flares</td>
<td>-15...-19</td>
<td>30 - 350</td>
<td>$10^{-6}$ Mpc$^{-3}$ yr$^{-1}$</td>
<td>6000</td>
</tr>
<tr>
<td>Luminous Red Novae</td>
<td>-9...-13</td>
<td>20 - 60</td>
<td>$10^{-13}$ yr$^{-1}$ Lsun$^{-1}$</td>
<td>80 - 3400</td>
</tr>
<tr>
<td>Fallback SNe</td>
<td>-4...-21</td>
<td>0.5 - 2</td>
<td>$&lt;5 \times 10^{-6}$ Mpc$^{-3}$ yr$^{-1}$</td>
<td>&lt; 800</td>
</tr>
<tr>
<td>SNe Ia</td>
<td>-17...-19.5</td>
<td>30 - 70</td>
<td>$3 \times 10^{-5}$ Mpc$^{-3}$ yr$^{-1}$</td>
<td>200000</td>
</tr>
<tr>
<td>SNe II</td>
<td>-15...-20</td>
<td>20 - 300</td>
<td>$(3.8) \times 10^{-5}$ Mpc$^{-3}$ yr$^{-1}$</td>
<td>100000</td>
</tr>
</tbody>
</table>

Table adapted from Rau et al. 2009 by Lucianne Walkowicz
NOAO’s proposed broker Antares

- Solar System
- LSST History
- Other catalogs
- Ancillary data

0.1 rare alerts/image

Saha et al
1409.0056
NOAO’s proposed broker Antares

- Solar System
- LSST History
- Other catalogs
- Ancillary data

0.1 rare alerts/image

2017 workshop
2016 LSST AHM

Saha et al
1409.0056
Bayesian Networks

Very broadly speaking 5 flavors of BNs possible
- Naïve
- Tree Augmented Network (TAN)
- **Constructed (semantics, expert knowledge etc. based)**
- Single winner from several naïve
- Fully learned from data

Search space growth hyperexponential

<table>
<thead>
<tr>
<th>n</th>
<th>G(n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>543</td>
</tr>
<tr>
<td>5</td>
<td>29,281</td>
</tr>
<tr>
<td>6</td>
<td>3,781,503</td>
</tr>
<tr>
<td>7</td>
<td>$1.1 \times 10^9$</td>
</tr>
<tr>
<td>8</td>
<td>$7.8 \times 10^{11}$</td>
</tr>
<tr>
<td>9</td>
<td>$1.2 \times 10^{15}$</td>
</tr>
<tr>
<td>10</td>
<td>$4.2 \times 10^{18}$</td>
</tr>
</tbody>
</table>
SNe/non-SNe BN

Based on peaks normalized

80-90% completeness
Only archival information

prior outburst = \frac{1}{t_{span}} \cdot \left( \sum_i w_i(p_i - p_m)^2 \right)^{1/2}
Discriminating features

You can not step into the same river twice.
Hierarchical approach

Archival search

Binary Blackholes
Graham et al. 2015
CARMA/Wavelets
Many features - not all are independent

Resort to dimensionality reduction
Challenge 3: A variety of parameters - choose judiciously

Discovery; Contextual; Follow-up; Prior Classification …

Whole curve measures
Median magnitude (mag); mean of absolute differences of successive observed magnitude; the maximum difference magnitudes

Fitted curve measures
Scaled total variation scaled by number of days of observation; range of fitted curve; maximum derivative in the fitted curve

Residual from fit measures
The maximum studentized residual; SD of residuals; skewness of residuals; Shapiro-Wilk statistic of residuals

Cluster measures
Fit the means within the groups (up to 4 measurements); and then take the logged SD of the residuals from this fit; the max absolute residuals from this fit; total variation of curve based on group means scaled by range of observation
Challenge 4: real-time computation required - find ways to make that happen

Recomputation of features

Updating priors

Ridgeway et al., arXiv: 1409.3265
Challenge 4: real-time computation required - find ways to make that happen

Recomputation of features

Updating priors

Ridgeway et al., arXiv: 1409.3265
Challenge 5: Metaclassification - combining diverse classifiers optimally

As varied classifiers are used for parts of the classification tree combing their outputs in an optimal way becomes crucial

Mahabal, Donalek
Sky Maps of a few surveys

CRTS

PTF

Gaia

LSST
Why Domain Adaptation?

• Surveys differ in depth (aperture), filters, cadence

• Same (type of) objects produce different statistical features (skew, median absolute deviation etc.)

• Learning tends to be done on each survey separately - leading to unnecessary delays

• DA helps build on the otherwise untapped intersurvey synergy (think DASCH -> CRTS/ZTF/Kepler -> LSST)

Jingling Li, S Vaijanapurkar, B Bue
Feature Correlations
KIC 8462852 (aka Tabby’s star aka WTF star)

Fading or not fading?

nir and UV flux
SMOTE and Sampling with replacement used to take care of unbalancedness

50K Variables from CRTS

Drake et al. 2014

Selected class distribution in CRTS

Selected class distribution in Lineardb

Selected class distribution in PTF(R)
If you had just two features

source domain

target domain

target domain

Feature 1
Feature 2

Feature 1
Feature 2

Feature 2

promote instances
Geodesik Flow Kernel

- Integrate flow of subspace: S to T
- Kernel encapsulates incremental changes between subspaces
- Kernel converts domain specific features into invariant ones (Gong et al. 2012)
Co-Domain Adaptation

- Slow adaptation from S to T
- Add best target objects in each round
- Elect shared S and T subsets from training and unlabelled data (Chen et al. 2011)

\[ L = D_S \cup D_T^l \]
\[ U = D_T^u \]

Adding a fraction of sources from the target domain to the source domain for training improves performance
Summary of challenges

- Characterize/Classify as much with as little data as possible
- Only a small fraction are rare - find/characterize them early
- A variety of parameters - choose judiciously
- Real-time computation is required - find ways to make that happen
- Metaclassification - combining diverse classifiers optimally
Summary of challenges

• Characterize/Classify as much with as little data as possible
• Only a small fraction are rare - find/characterize them early
• A variety of parameters - choose judiciously
• Real-time computation is required - find ways to make that happen
• Meta-classification - combining diverse classifiers optimally

These challenges involve:

• Making sense of unparalleled volumes of structured and unstructured data in real-time, and
• Teaching machines how humans think by understanding pattern recognition when handling diverse types of data sources
Summary of challenges

- Characterize/Classify as much with as little data as possible
- Only a small fraction are rare - find/characterize them early
- A variety of parameters - choose judiciously
- Real-time computation is required - find ways to make that happen
- Metaclassification - combining diverse classifiers optimally

These challenges involve:
- Making sense of unparalleled volumes of structured and unstructured data in real-time, and
- Teaching machines how humans think by understanding pattern recognition when handling diverse types of data sources

Better tools to make sense of very sparse data and Streamlined workflows