Astronomical Applications of Machine Learning and Neural Networks

Ashish Mahabal <aam@astro.caltech.edu>
Center for Data Driven Discovery, Caltech
LISA, KISS, Caltech, 20180117
Figure 4. Framework of Cosmic Explosions in the Year 2011 (Kasliwal 2011). Note that until 2005 (Fig. 1), we only knew about three classes (denoted by gray bands). In the past six years, systematic searches, serendipitous discoveries and archival searches have uncovered a plethora of novel, rare transients. Discoveries by the Palomar Transient Factory and P60-FasTING (Kasliwal et al. 2011a) are denoted by .Ia Explosions (helium detonations in ultra-compact white dwarf binaries), pair instability explosions, supernovae (magnetars), luminous red novae (electron capture induced collapse of rapidly rotating O–Ne–Mg white dwarfs), luminous supernovae (magnetars), new classes are emerging and the governing physics is being widely debated: luminous red novae (electron capture induced collapse of rapidly rotating O–Ne–Mg white dwarfs). Therefore, prior to the ambitious search for an electromagnetic counterpart to a gravitational wave signal, it would only be prudent to build this capability between gravitational wave searches and the electromagnetic search described above is that localizations of the gravitational wave signal and consequent large false positive rate of electromagnetic searches is expected to become routine. A basic commonality between gravitational wave searches and the electromagnetic search described above is that waves from neutron star mergers every month is expected to become routine. A basic commonality between gravitational wave searches and the electromagnetic search described above is that waves from neutron star mergers every month is expected to become routine.

Several. .Ia Explosions (helium detonations in ultra-compact white dwarf binaries), pair instability explosions, supernovae (magnetars), luminous red novae (electron capture induced collapse of rapidly rotating O–Ne–Mg white dwarfs), luminous supernovae (magnetars), new classes are emerging and the governing physics is being widely debated: luminous red novae (electron capture induced collapse of rapidly rotating O–Ne–Mg white dwarfs). Therefore, prior to the ambitious search for an electromagnetic counterpart to a gravitational wave signal, it would only be prudent to build this capability between gravitational wave searches and the electromagnetic search described above is that localizations of the gravitational wave signal and consequent large false positive rate of electromagnetic searches is expected to become routine. A basic commonality between gravitational wave searches and the electromagnetic search described above is that waves from neutron star mergers every month is expected to become routine.
Overwhelming (amounts of) data

500 Million Light Curves with ~ $10^{11}$ data points

- RR Lyrae
- W Uma
- Eclipsing
- Flare star (UV Ceti)
- CV
- Blazar

ZTF, LSST, SKA
Properties of light-curves

- Gappy
- Irregular
- Heteroskedastic

Reasons:

- expense, rotation/revolution of Earth, moon
- science objectives, weather, moon
- weather, moon, airmass

errors ignored by many methods
Statistical features

Compute features (statistical measures) for each light curve: amplitudes, moments, periodicity, etc.
Converts heterogeneous light curves into homogeneous **feature vectors** in the parameter space
Apply a variety of automated classification methods

![Graph showing statistical features](image)
Light-curve features

**Graph:**
- **V mag** on the y-axis.
- **Phase** on the x-axis.
- **Normal RR Lyrae** as indicated in the graph.
- **Period** highlighted.
- **Amplitude** highlighted.
- **Mean** highlighted.
- **Median** highlighted.

**Data:**
- Median: (value)
- Mean: (value)
- Amplitude: (value)
- Period: (value)
### Statistical features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>meanmag</td>
<td>(&lt; \text{mag} &gt;)</td>
</tr>
<tr>
<td>minmag</td>
<td>(\text{mag}_{\text{min}})</td>
</tr>
<tr>
<td>maxmag</td>
<td>(\text{mag}_{\text{max}})</td>
</tr>
<tr>
<td>amplitude</td>
<td>(0.5 \times (\text{mag}<em>{\text{max}} - \text{mag}</em>{\text{min}}))</td>
</tr>
<tr>
<td>beyond 1 std</td>
<td>(p(|\text{mag} - \langle \text{mag} \rangle| &gt; \sigma))</td>
</tr>
<tr>
<td>flux percentile ratio mid 20</td>
<td>(\frac{(\text{flux}<em>{60} - \text{flux}</em>{40})}{(\text{flux}<em>{95} - \text{flux}</em>{5})})</td>
</tr>
<tr>
<td>flux percentile ratio mid 35</td>
<td>(\frac{(\text{flux}<em>{67.5} - \text{flux}</em>{32.5})}{(\text{flux}<em>{95} - \text{flux}</em>{5})})</td>
</tr>
<tr>
<td>flux percentile ratio mid 50</td>
<td>(\frac{(\text{flux}<em>{75} - \text{flux}</em>{25})}{(\text{flux}<em>{95} - \text{flux}</em>{5})})</td>
</tr>
<tr>
<td>flux percentile ratio mid 65</td>
<td>(\frac{(\text{flux}<em>{82.5} - \text{flux}</em>{17.5})}{(\text{flux}<em>{95} - \text{flux}</em>{5})})</td>
</tr>
<tr>
<td>flux percentile ratio mid 80</td>
<td>(\frac{(\text{flux}<em>{90} - \text{flux}</em>{10})}{(\text{flux}<em>{95} - \text{flux}</em>{5})})</td>
</tr>
<tr>
<td>linear trend</td>
<td>(b) where (\text{mag} = a \times t + b)</td>
</tr>
<tr>
<td>max slope</td>
<td>(\max(|\text{mag}_{i+1} - \text{mag}<em>i|/(t</em>{i+1} - t_i)))</td>
</tr>
<tr>
<td>median absolute deviation</td>
<td>(\text{med}(\text{flux} - \text{flux}_{\text{med}}))</td>
</tr>
<tr>
<td>median buffer range percentage</td>
<td>(p(|\text{flux} - \text{flux}<em>{\text{med}}| &lt; 0.1 \times \text{flux}</em>{\text{med}}))</td>
</tr>
<tr>
<td>pair slope trend</td>
<td>(p(\text{flux}_{i+1} - \text{flux}_i &gt; 0; i = n - 30, n))</td>
</tr>
<tr>
<td>percent difference flux percentile skew</td>
<td>(\frac{\mu_3}{\sigma^3})</td>
</tr>
<tr>
<td>small kurtosis</td>
<td>(\frac{\mu_4}{\sigma^4})</td>
</tr>
<tr>
<td>std</td>
<td>(\sigma)</td>
</tr>
<tr>
<td>stetson j</td>
<td>(\text{var}_j(\text{mag}))</td>
</tr>
<tr>
<td>stetson k</td>
<td>(\text{var}_k(\text{mag}))</td>
</tr>
</tbody>
</table>
Many features
- not all are independent

Ashish Mahabal
15 Jan 2015

Adam Miller
20
# Feature selection strategy

- Fast Relief Algorithm (wt and threshold)
- Fisher Discriminant Ratio
- Correlation based Feature Selection
- Fast Correlation Based Filter
- Multi Class Feature Selection

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>pp</td>
<td>pp</td>
</tr>
<tr>
<td>data_num</td>
<td>data_num</td>
</tr>
<tr>
<td>amplitude</td>
<td>amplitude</td>
</tr>
<tr>
<td>linear_trend</td>
<td>linear_trend</td>
</tr>
<tr>
<td>ls</td>
<td>ls</td>
</tr>
<tr>
<td>mod_bsr_range_par</td>
<td>mod_bsr_range_par</td>
</tr>
<tr>
<td>std</td>
<td>std</td>
</tr>
<tr>
<td>pdfp</td>
<td>pdfp</td>
</tr>
<tr>
<td>percent_amplitude</td>
<td>percent_amplitude</td>
</tr>
<tr>
<td>max_slope</td>
<td>max_slope</td>
</tr>
<tr>
<td>skew</td>
<td>skew</td>
</tr>
<tr>
<td>kurtosis</td>
<td>kurtosis</td>
</tr>
<tr>
<td>tpr_md35</td>
<td>tpr_md35</td>
</tr>
<tr>
<td>tpr_md65</td>
<td>tpr_md65</td>
</tr>
<tr>
<td>tpr_md60</td>
<td>tpr_md60</td>
</tr>
<tr>
<td>tpr_md20</td>
<td>tpr_md20</td>
</tr>
<tr>
<td>ng</td>
<td>ng</td>
</tr>
<tr>
<td>mod_abs_dev</td>
<td>mod_abs_dev</td>
</tr>
<tr>
<td>tpr_md80</td>
<td>tpr_md80</td>
</tr>
<tr>
<td>dec</td>
<td>dec</td>
</tr>
<tr>
<td>beyond1std</td>
<td>beyond1std</td>
</tr>
<tr>
<td>ra</td>
<td>ra</td>
</tr>
<tr>
<td>pair_slope_trend</td>
<td>pair_slope_trend</td>
</tr>
<tr>
<td>magratio</td>
<td>magratio</td>
</tr>
<tr>
<td>rcorbor</td>
<td>rcorbor</td>
</tr>
</tbody>
</table>

Donalek, .., Mahabal, … arxiv:1310.1976
Real-Bogus classification
Citizen Science (zooniverse) + random forests
Umaa Rebbapragada
Promise: Works better

Pitfall: Blacker box

Adil Moujahid
Reduce breadth

Reduce depth

Go example creation: Bob van den Hoek

http://gobase.org/online/intergo/?query=%22hane%20nobi%22

non-image deep networks

Recurrent Neural Networks

one to one
one to many
many to one
many to many

http://colah.github.io/posts/2015-08-Understanding-LSTMs/
Long Short Term Memory (LSTM)

Can be used for light-curves and other time-series
Convolutional network (single slide) primer
Visualizations of Layer 1 and 2. Each layer illustrates 2 pictures, one which shows the filters themselves and one that shows what part of the image are most strongly activated by the given filter. For example, in the space labeled Layer 2, we have representations of the 16 different filters (on the left).

Promise: New Features

Ashish Ma  Visualizations of Layers 3, 4, and 5

Transient hunting through image differencing

reorient, match background, match PSF, eliminate artifacts
Image subtraction for hunting transients without subtraction

Sedaghat and Mahabal, 2017
50K Periodic Variables from CRTS

Distribution of all classes in CRTS

Selected class distribution in CRTS

Drake et al. 2014
(dmdt) Image representation

Area equalized pixels

light curve with n points

23 x 24 output grid

n * (n-1)/2 points

Mahabal et al., 2017
Network architecture

In this case, shallow works well too.
Random Forest using standard features

no features
no dimensionality reduction
comparable results

Convolutional Network

Binary probabilities are better
\[ dmdt \text{–image} = b + ci + s \]

- background (survey, cadence)
- class background
- individual object (specific)

\[
\begin{align*}
\text{Min} & \left\| M - L - S \right\|_2 \\
L, S & \\
1. & L \text{ lies in the set of low-rank matrices}, \\
2. & S \text{ lies in the set of sparse matrices}.
\end{align*}
\]

non-convex robust PCA
Netrapalli et al., 2014
Each class is like a different road
Each individual object has/is perturbations over it

Andrew Kirillov

Video Surveillance Anology

non-convex robust PCA
Netrapalli et al., 2014
EW/EB separation?

Two separate backgrounds emerged for class 1.
LIGO and VIRGO have announced our first Binary Neutron Star gravitational-wave event! Look for a special surprise when classifying on workflow Neutron Star Merger and above. Facts about the event can be found here: http://www.ligo.org/detections/GW170817.php and papers about the discovery can be found here: https://www.ligo.caltech.edu/page/detection-companion-papers
FIG. 1. Sample signal injected into real LIGO noise. The red time-series is an example of the input to our Deep Filtering algorithm. It contains a hidden BBH GW signal (blue) from our test set which was superimposed in real LIGO noise from the test set and whitened. For this injection, the optimal matched-filter SNR = 7.5 (peak power of this signal is 0.65 times the power of background noise). The component masses of the merging BHs are 57\(M_\odot\) and 33\(M_\odot\). The presence of this signal was detected directly from the (red) time-series input with over 99% sensitivity and the source’s parameters were estimated with a mean relative error less than 10%.

FIG. 2. Spectrograms of real LIGO noise test samples. We used signals injected into real data from the LIGO detectors in this article, ensuring that the training and testing sets did not contain noise from the same events. These are some random examples of real glitches that were present in our test set of LIGO noise. The Deep Filtering method takes the 1D strain directly as input and is able to correctly classify glitches as noise and detect true GW signals as well as simulated GW signals injected into these highly non-stationary non-Gaussian data streams, with similar sensitivity compared to matched-filtering.
LIGO auxiliary channels

Snow plows
Earthquakes
High Winds

different frequencies/binning
different time signatures

Characterize and identify
in streaming data

With Jess McIver
Other astro applications

- Supernova classification, Charnock and Moss, 1606.07442
- Supernova real-bogus, Cabrera-Vives et al., 1701.00458
- Star-galaxy separation, Kim and brunner, 1608.04369
- Radio galaxies, Aniyan and Thorat, 1705.03413
- Galaxy bars, Abraham et al., 1711.04573
Summary

• Direct light curve classification using DL
• Adapting to other surveys
• Applicability to radio, x-ray etc.
• Applicability to transients (with sparse lc)

Extension to other forms e.g. spectra possible

Plans to apply to gravitational wave data

Deep learning is here to stay!

aam@astro.caltech.edu