Astroinformatics in the Time Domain:

Classification of Light Curves and Transients

Prof. S. George Djorgovski

With: M. Graham, A. Mahabal, A. Drake,

and many students and collaborators

Center for Data-Driven Discovery and Astronomy Dept., Caltech

Lecture 3

XXX Canary Islands Winter School

November 2018
What can we observe?

Astronomy in SpaceTime

Traditional astronomy is on the 3D hyper-surface (aka space) of the past light cone in the 4D spacetime.

Time-domain astronomy carves out a 4D hyper-volume as we move along the time axis of the 4D spacetime.
Astronomy in the Time Domain

- Rich phenomenology, from the Solar system to cosmology and extreme relativistic physics
  - Touches essentially every field of astronomy
- For some phenomena, time domain information is a key to the physical understanding
- A qualitative change:
  - Static $\Rightarrow$ Dynamic sky
  - Sources $\Rightarrow$ Events
- Real-time discovery/reaction requirements pose new challenges for knowledge discovery

Synoptic, panoramic surveys $\Rightarrow$ event discovery
Rapid follow-up and multi-$\lambda$ $\Rightarrow$ keys to understanding
Synoptic Sky Surveys

- Synoptic digital sky surveys – i.e., a panoramic cosmic cinematography – are now the dominant data producers in astronomy
  - From Terascale to Petascale data streams
- A major new growth area of astrophysics
  - Driven by the new generation of large digital synoptic sky surveys (CRTS, PTF/ZTF, PanSTARRS, SkyMapper, ...), leading to LSST, SKA, etc.
- A broader significance for an automated, real-time knowledge discovery in massive data streams
Characterizing Synoptic Sky Surveys

Define a measure of depth (roughly \( \sim \) S/N of indiv. exposures):

\[
D = \left[ A \times t_{\text{exp}} \times \varepsilon \right]^{1/2} / \text{FWHM}
\]

where
- \( A \) = the effective collecting area of the telescope in \( \text{m}^2 \)
- \( t_{\text{exp}} \) = typical exposure length
- \( \varepsilon \) = the overall throughput efficiency of the telescope+instrument
- \( \text{FWHM} \) = seeing

Define the **Scientific Discovery Potential** for a survey:

\[
SDP = D \times \Omega_{\text{tot}} \times N_b \times N_{\text{avg}}
\]

where
- \( \Omega_{\text{tot}} \) = total survey area covered
- \( N_b \) = number of bandpasses or spec. resolution elements
- \( N_{\text{avg}} \) = average number of exposures per pointing

**Transient Discovery Rate:**

\[
TDR = D \times R \times N_e
\]

where
- \( R = d\Omega/dt \) = area coverage rate
- \( N_e \) = number of passes per night
Parameter Spaces for the Time Domain

(in addition to everything else: flux, wavelength, etc.)

• For *surveys*:
  o Total exposure per pointing
  o Number of exposures per pointing
  o How to characterize the cadence?
    ➔ Window function(s)
    ➔ Inevitable biases

• For *objects/events* ~ light curves:
  o Significance of periodicity, periods
  o Descriptors of the power spectrum (e.g., power law)
  o Amplitudes and their statistical descriptors
  ... etc. – over 70 parameters defined so far, but which ones are the minimum / optimal set?
The Palomar-Quest Event Factory


Real-time detection and publishing of transients using \textit{VOEvent}

- Precursor of the PTF
- Progenitor of the CRTS

Young SNe Ia, P200 spectra $\sim 1$h after the initial detection
Automating Real-Time Astronomy

- Cyber-infrastructure for time domain astronomy
- VOEvent standard for real-time publishing/requests
- VOEventNet: A telescope network with a feedback
- Scientific measurements spawning other measurements and data analysis in the real time

PI: R. Williams

Now skyalert.org
The Transient Alert Data Environment

"Here's an event"
"Send me everything matching these criteria"
"Here's what I learned"
"Observe X with parameters Y"
"Tell me status of X"
"Send me data of X"

R. Street, LCO
Catalina Real-Time Transient Survey (CRTS)

• Data from a search for near-Earth asteroids at UA/LPL; we discover astrophysical transients in their data stream
• 3 (now 2) telescopes in AZ, AU
• > 80% of the sky covered ~ 300 – 500 times down to ~ 19 – 21 mag, baselines 10 min to 12 yrs
• So far ~ **17,000 transients**, including > 4,000 SNe, > 1,500 CVs, ~ 5,000 AGN, etc.

**Open data policy:** all data are made public; transients are published immediately on line, for the entire community
A Variety of CRTS Transients

SNe

Blazars/AGN

GRB afterglows

CVs

Flare stars

Eclipses and occultations
Event Publishing / Dissemination

- Real time: VOEvent, RSS, (initially also SkyAlert, Twitter, iApp)
- Next day: annotated tables on the CRTS website

<table>
<thead>
<tr>
<th>CRTS 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>FC</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSS140107:082255+073033</td>
<td>125.72901</td>
<td>7.50903</td>
<td>20140107</td>
<td>18.88</td>
<td>1401071070454146201</td>
<td>yes</td>
<td>14620</td>
<td>no</td>
<td>2014-01-07</td>
<td>14620</td>
<td>yes</td>
<td>Unknown SDSS mag 22</td>
</tr>
<tr>
<td>CSS140107:015854+053524</td>
<td>29.72348</td>
<td>5.59009</td>
<td>20140107</td>
<td>15.33</td>
<td>140107104114136538</td>
<td>yes</td>
<td>13653</td>
<td>no</td>
<td>2014-01-07</td>
<td>13653</td>
<td>yes</td>
<td>SN 2013hs (Howerton) mag 1</td>
</tr>
<tr>
<td>CSS140107:114807+014254</td>
<td>177.02819</td>
<td>1.71506</td>
<td>20140107</td>
<td>19.34</td>
<td>1401071010634128518</td>
<td>yes</td>
<td>12851</td>
<td>no</td>
<td>2014-01-07</td>
<td>12851</td>
<td>yes</td>
<td>QSO SDSS mag 20,5</td>
</tr>
<tr>
<td>CSS140107:145029-083859</td>
<td>222.61934</td>
<td>-8.64971</td>
<td>20140107</td>
<td>14.16</td>
<td>1401070909794151988</td>
<td>yes</td>
<td>15198</td>
<td>no</td>
<td>2014-01-07</td>
<td>15198</td>
<td>yes</td>
<td>HPM LHS_381</td>
</tr>
<tr>
<td>CSS140107:230108-015805</td>
<td>345.28333</td>
<td>-1.96798</td>
<td>20140107</td>
<td>17.75</td>
<td>1401070901123412940</td>
<td>yes</td>
<td>11294</td>
<td>no</td>
<td>2014-01-07</td>
<td>11294</td>
<td>yes</td>
<td>Blazar mag 19,0</td>
</tr>
<tr>
<td>CSS140104:013741+220312</td>
<td>24.42141</td>
<td>22.05338</td>
<td>20140104</td>
<td>14.51</td>
<td>1401041210049138406</td>
<td>yes</td>
<td>13840</td>
<td>no</td>
<td>2014-01-04</td>
<td>13840</td>
<td>yes</td>
<td>Unknown SDSS mag 22</td>
</tr>
<tr>
<td>CSS140104:020225+144325</td>
<td>30.60214</td>
<td>14.72357</td>
<td>20140104</td>
<td>19.89</td>
<td>1401041150114109714</td>
<td>yes</td>
<td>10971</td>
<td>no</td>
<td>2014-01-04</td>
<td>10971</td>
<td>yes</td>
<td>AGN SDSS mag 21,9</td>
</tr>
<tr>
<td>CSS140104:034718+014254</td>
<td>56.82594</td>
<td>1.71497</td>
<td>20140104</td>
<td>19.42</td>
<td>140104101021427020</td>
<td>no</td>
<td>12702</td>
<td>no</td>
<td>2014-01-04</td>
<td>12702</td>
<td>yes</td>
<td>SN mag 16</td>
</tr>
</tbody>
</table>

**Discovery data**

**Archival data**

**Light curve+images**

**Finding chart**
500 Million Light Curves with $>10^{11}$ data points

- **RR Lyrae**
- **W Uma**
- **Eclipsing**
- **Flare star (UV Ceti)**
- **CV**
- **Blazar**
Zwicky Transient Facility (2017-)

- New camera on Palomar Oschin 48” with 47 deg$^2$ field of view
- 3750 deg$^2$/hr to 20.5-21 mag (1.2 TB/night)
- Full northern sky (~12,000 deg$^2$) every three nights
- Galactic Plane every night
- Over 3 years: 3 PB, 750 billion detections, ~1000 detections/src
- First megaevent survey: $10^6$ alerts per night (Apr 2018)
# ZTF = 0.1 LSST

<table>
<thead>
<tr>
<th></th>
<th>ZTF</th>
<th>LSST</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of sources</td>
<td>1 billion</td>
<td>37 billion</td>
</tr>
<tr>
<td>No. of detections</td>
<td>1 trillion</td>
<td>37 trillion</td>
</tr>
<tr>
<td>Annual visits per source</td>
<td>1000 (2+1 filters)</td>
<td>100 (6 filters)</td>
</tr>
<tr>
<td>No. of pixels</td>
<td>600 million (1320 cm² CCDs)</td>
<td>3.2 billion (3200 cm² CCDs)</td>
</tr>
<tr>
<td>Field of view</td>
<td>47 deg²</td>
<td>9 deg²</td>
</tr>
<tr>
<td>Hourly survey rate</td>
<td>3750 deg²</td>
<td>1000 deg²</td>
</tr>
<tr>
<td>Nightly alert rate</td>
<td>1 million</td>
<td>10 million</td>
</tr>
<tr>
<td>Nightly data rate</td>
<td>1.4 TB</td>
<td>15 TB</td>
</tr>
</tbody>
</table>
Automated Classification of Transients

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

Rapid, automated transient classification is a critical need!
Event Classification is a Hard Problem

- Classification of transient events is essential for their astrophysical interpretation and uses
  - Must be done in real time and iterated dynamically
- Human classification is already unsustainable, and will not scale to the Petascale data streams
- This is hard:
  - Data are sparse and heterogeneous: feature vector approaches do not work; using Bayesian approach
  - Completeness vs. contamination 🐐
  - Follow-up resources are expensive and/or limited: only the most interesting events
  - Iterate classifications dynamically as new data come in
- Traditional DP pipelines do not capture a lot of the relevant contextual information, prior/expert knowledge, etc.
Spectroscopic Follow-up is a Critical Problem (and it will get a lot worse)

- Recently: data streams of $\sim 0.1 \, \text{TB} / \text{night}$, $\sim 10^2$ transients / night (CRTS, PTF, various SN surveys, microlensing, etc.)
  - We were already in the regime where we cannot follow them all
  - Spectroscopy is the key bottleneck now, and it will get worse
- Now (ZTF): $\sim 1 \, \text{TB} / \text{night}$, $\sim 10^5 - 10^6$ transients / night (PanSTARRS, Skymapper, VISTA, VST, SKA precursors…)
- Forthcoming (soonish?): LSST, $\sim 30 \, \text{TB} / \text{night}$, $\sim 10^7$ transients / night, SKA
- So… which ones will you follow up?
- Follow-up resources will likely remain limited

A major, qualitative change!

Transient classification is essential
Towards an Automated Event Classification

- Incorporation of the contextual information (archival, and from the data themselves) is essential
- Automated prioritization of follow-up observations, given the available resources and their cost
- A dynamical, iterative system

**Flowchart:**

1. Input Data and Event Streams
2. Evolving Event Portfolios
3. Event Classification Engine
4. Evolving Classification Probabilities: $P(SN\ Ia) = \ldots$, $P(SN\ II) = \ldots$, $P(AGN) = \ldots$, $P(CV) = \ldots$, $P(dM) = \ldots$ ...
5. Follow-up Facilities
6. Follow-up Prioritization & Decision Engine
Automated Detection of Artifacts

Automated classification and rejection (> 95%) of artifacts masquerading as transient events in the PQ survey pipeline, using a Multi-Layer Perceptron ANN

(C. Donalek)
A Variety of Classification Methods

- **Bayesian Networks**
  - Can incorporate heterogeneous and/or missing data
  - Can incorporate contextual data, e.g., distance to the nearest star or galaxy

- **Probabilistic Structure Functions**
  - A new method, based on 2D \([\Delta t, \Delta m]\) distributions
  - Now expanding to data point triplets: \(\Delta t_{12}, \Delta m_{12}, \Delta t_{23}, \Delta m_{23}\), giving a 4D histogram

- **Random Forests**
  - Ensembles of Decision Trees

- **Feature Selection Strategies**
  - Optimizing classifiers

- **Machine-Assisted Discovery**
  - etc., etc.
A Hierarchical Approach to Classification

Different types of classifiers perform better for some event classes than for the others.

We use some astrophysically motivated major features to separate different groups of classes.

Proceeding down the classification hierarchy, every node uses those classifiers that work best for that particular task.
Data are Sparse and Heterogeneous

Bayesian approaches

Generating priors for various observables for different types of variables

(Lead: A. Mahabal)
Gaussian Process Regression (GPR)

A generalization of a Gaussian probability, specified by a mean function and a positive definite covariance function.

Given two flux measurement points for a new transient we can ask which of the different models it fits, and what stage of their period or phase. The more points you have, the better the estimate.
2D Light Curve Priors

• For any pair of light curve measurements, compute the $\Delta t$ and $\Delta m$, make a 2D histogram
  – $N$ independent measurements generate $N^2$ correlated data points

• Compare with the priors for different types of transients

• Repeat as more measurements are obtained, for an evolving, constantly improving classification

• Now expanding to consecutive data point triplets: $\Delta t_{12}, \Delta m_{12}, \Delta t_{23}, \Delta m_{23}$, giving a 4D histogram

(Lead: B. Moghaddam)
Applying $\Delta m$ vs. $\Delta t$ Histograms

- Measure of a divergence between the unknown transient histogram and two prototype class histograms.
Δm vs. Δt Classifier Performance

- Performance measured using Leave-one-out cross-validation (LOOCV)

<table>
<thead>
<tr>
<th></th>
<th>SN</th>
<th>CVBlazarRRMira</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN</td>
<td>A0 = 96.5%</td>
<td>3.5%</td>
</tr>
<tr>
<td>CVBlazarRRMira</td>
<td>2.1%</td>
<td>A1 = 97.9%</td>
</tr>
</tbody>
</table>

- Optimize histogram parameters (binning, smoothing, Dirichlet prior parameters) using a genetic algorithm

- A modest, but a consistent improvement over the human expert selected parameters (Y. Chen, C. Donalek)
A New Approach Using Convolutional ANN

A. Mahabal et al. 2017, IEEE Computational Intelligence 2017, p. 2757 = arxiv/1709.06257
From Light Curves to Feature Vectors

• We compute ~ 70 parameters and statistical measures for each light curve: amplitudes, moments, periodicity, etc.
• This turns *heterogeneous* light curves into *homogeneous feature vectors* in the parameter space
• Apply a variety of automated classification methods
Variability Feature Space

• Generate homogeneous representation of time series
• Most Richards et al. (2011) features carry little information
• Measuring:
  – Morphology (shape): skew, kurtosis
  – Scale: Median absolute deviation, biweight midvar.
  – Variability: Stetson, Abbe, von Neumann
  – Timescale: periodicity, coherence, characteristic
  – Trends: Thiel-Sen
  – Autocorrelation: Durbin-Watson
  – Long-term memory: Hurst exponent
  – Nonlinearity: Teraesvirta
  – Chaos: Lyapunov exponent
  – Models: HMM, CAR, Fourier decomposition, wavelets
• Defines high-dimensional (representative) feature space
Automated Classification of Variable Stars

Dubath et al. (2011): Used random forests on a set of 14 light curve features to recover 26 classes of variable stars from the Hipparcos catalog.

Confusion matrix ==> Similar results by the Berkeley group (Richards et al. 2011)
Light Curves Clustering in Feature Space

- Unsupervised Machine Learning
- Can be used to determine the number of classes and cluster the input data in classes on the basis of their statistical properties only
- Search for Outliers, Trajectories, etc.
- Methods: SOM, K-means, Hierarchical Clustering, etc.
- Given a set of features, which ones are the most discriminating between different classes?
Principal Component Analysis (PCA)

Solving the eigen-problem of the data hyperellipsoid in the parameter space of measured attributes

$$p_1 = \text{observables} \quad (i = 1, \ldots D_{\text{data}})$$

$$\xi_j = \text{eigenvectors, or principal axes of the data hyperellipsoid}$$

$$e_j = \text{eigenvalues, or amplitudes of } \xi_j \quad (j = 1, \ldots D_{\text{stat}})$$
Correlation Searches in Attribute Space

Data dimension $D_D = 2$
Statistical dim. $D_S = 2$

If $D_S < D_D$, correlations are present

Correlations are clusters with dimensionality reduction

A real-life example:
“Fundamental Plane” of elliptical galaxies, a set of bivariate scaling relations in a parameter space of $\sim 10$ dimensions, containing valuable insights into their physics and evolution.
What About the Clustering?
Feature Selection Algorithms

They are a subset of **dimensionality reduction** techniques.

- **Filter methods** apply a statistical measure to assign a scoring to each feature, usually independently (univariate). The features are ranked by the score.

- **Wrapper methods** look for a set of features where different feature combinations are evaluated and compared to other combinations.

- **Embedded methods** learn which features best contribute to the accuracy of the model while the model is being created.

- The **scoring criterion** depends on the goal, e.g.:
  - Accurate predictions for the regression searches
  - Classification discrimination power for clustering
Feature Selection Algorithms: Examples

• **Fast Relief Algorithm** (aka ReliefF) ranks features according to how well their values distinguish between instances.

• **Fisher Discriminant Ratio** (FDR) ranks features according to their classification discriminatory power. It can be applied only to binary classification problems.

• **Correlation-based Feature Selection** (CFS) is a wrapper method which selects features that have low redundancy (i.e., not correlated with each other) and is strongly predictive of a class.

• **Fast Correlation Based Filter** (FCBF) is a supervised filter algorithm, similar to the CFS. Searches for features that have predominant correlation with the class. Can be computationally efficient with very high dimensional data.

• **Multi Class Feature Selection** (MCFS) is an unsupervised method based on the spectral analysis of the data. ... etc.
Feature Selection Algorithms

Optimal sets of features may be different for

• Different regression target variables:
  e.g., \( y_1 = f_1(x_i, x_j, x_k, ...) \), \( y_2 = f_2(x_p, x_q, x_r, ...) \), etc.

• Different classification tasks:
  e.g., \( \text{Class (A,B)} = f(x_a, x_b, x_c, ...) \), \( \text{Class (A,B,C)} = f(x_d, x_e, x_f, ...) \)

• Different regression or classification algorithms:
  e.g., ANN, DT, RF, SVM, ...

  . . . so they have to be optimized in each individual case

See:

Donalek et al., IEEE BigData 2013, p. 35 = arxiv/1310.1976
Optimizing Feature Selection

Select a subset of features from the data matrix $X$ that best predict the data in classes $Y$ by sequentially selecting features until there is no improvement in prediction: using Decision Trees with a 10-fold cross validation.

<table>
<thead>
<tr>
<th></th>
<th>Completeness</th>
<th>Contamination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blazar</td>
<td>83%</td>
<td>13%</td>
</tr>
<tr>
<td>CV</td>
<td>94%</td>
<td>6%</td>
</tr>
<tr>
<td>RR Lyrae</td>
<td>97%</td>
<td>4%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Completeness</th>
<th>Contamination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blazar</td>
<td>81%</td>
<td>13%</td>
</tr>
<tr>
<td>CV</td>
<td>96%</td>
<td>5%</td>
</tr>
<tr>
<td>SN Ia</td>
<td>100%</td>
<td>&lt;1%</td>
</tr>
</tbody>
</table>

Amplitude beyond 1std \(\text{flux\_percentile\_ratio\_mid65}\) max_slope qso std lomb-scargle

Linear_trend Median_absolute_deviation lomb-scargle

(Lead: C. Donalek)
Rank features in the order of classification quality for a given classification problem, e.g., RR Lyrae vs. WUMa

(Lead: C. Donalek)
Contextual Information is Essential

- **Visual context** contains valuable information about the reality and classification of transients.
- So does the **temporal context**, from the archival light curves.
- And the **multi-wavelength context**.
- Initial detection data contain little information about the transient: $\alpha$, $\delta$, $m$, $\Delta m$, ($t_c$). *Almost all of the initial information is archival or contextual;* follow-up information trickles in slowly, if at all.
- The importance and role of the archival information can only grow.
Bayesian Networks (BN): An Example

- Use the available measurements, missing data are not an issue
- Can use heterogeneous data, e.g., colors, flux changes, proximity to the nearest star or a galaxy (in projection)

\[ x = \text{input measurements of individual kinds (e.g., mags, colors, etc.)} \]
\[ y = \text{classes of events, } y = 1, \ldots, k. \]

Then:

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

Initial results for Supernova vs. non-Supernova classification, using a 3 parameter network:

Completeness \(\sim 80 - 90\%\)
Contamination \(\sim 10 - 20\%\)

Can be improved with the additional observables

(Lead: A. Mahabal)
Bayesian Networks: Implementation

*(Lead: A. Mahabal)*

Can incorporate contextual parameters, e.g., the normalized distances to the nearest star and the nearest galaxy as one of the BN variables.

\[\Rightarrow\]

\(\Rightarrow\) Rank light curve features in the order of the classification discrimination power.
Machine Discovery of Relationships

(see Graham et al. 2013, MNRAS 431, 2371 )

• Employs **symbolic regression** to determine best-fitting functional form to data and its parameters simultaneously

• Specify building blocks to be used: algebraic operators, analytical functions, constants

• Test: rediscover known astrophysical correlations (HRD, FP)

• An experiment in a binary classification of variable stars:
  o Characterize with \(~70\) periodic/non-periodic features
  o Use *Eureqa* for binary classification: *class 1* vs. *class 2*
  o Fit:  
  \[
  \text{class} = \text{step}[f(x_1, x_2, x_3, \ldots, x_{60})]
  \]
Classifying Light Curves with *Eureqa*

Light curves of two known stellar classes:

- Eclipsing binary (W U Ma)
- Pulsating variable (RR Lyrae)

Test using independent features

<table>
<thead>
<tr>
<th>Data set</th>
<th>Eureqa™ Purity</th>
<th>Eureqa™ Efficiency</th>
<th>Decision Tree Purity</th>
<th>Decision Tree Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR Lyrae</td>
<td>98%</td>
<td>96%</td>
<td>95%</td>
<td>95%</td>
</tr>
<tr>
<td>W UMa</td>
<td>97%</td>
<td>99%</td>
<td>96%</td>
<td>96%</td>
</tr>
<tr>
<td>CV</td>
<td>89%</td>
<td>91%</td>
<td>92%</td>
<td>92%</td>
</tr>
<tr>
<td>Blazar</td>
<td>68%</td>
<td>63%</td>
<td>87%</td>
<td>83%</td>
</tr>
<tr>
<td>SN Ia</td>
<td>76%</td>
<td>93%</td>
<td>90%</td>
<td>96%</td>
</tr>
<tr>
<td>CC SN</td>
<td>74%</td>
<td>41%</td>
<td>92%</td>
<td>80%</td>
</tr>
</tbody>
</table>
**Metaclassification:** An optimal combining of classifiers

Exploring a variety of techniques for an optimal classification fusion:
Markov Logic Networks, Diffusion Maps, Multi-Arm Bandit, Sleeping Expert…
For the *potentially most interesting events*, what type of follow-up observations has the greatest potential to discriminate among the competing event classes, given the available assets, and the potential scientific value?
Automating the Optimal Follow-Up

For the *potentially most interesting events*, what type of follow-up observations a $x$ has the greatest potential to discriminate among the competing event classes $y$, given the available assets, and the potential scientific value?

Request the optimal follow-up observations from the available assets that maximize the entropy drop:

$$H[p(y | x_+, x_0)] = - \sum_{y,x_+} p(y, x_+ | x_0) \log p(y | x_+, x_0)$$
Some Closing Thoughts

- Time domain astronomy requires an interconnected ecosystem of survey and follow-up telescopes, archives, and computational assets, which we do not yet have
  - Coordinated complementary time cadences
  - Multi-λ co-observing
- Transients (time-critical events) may be becoming less interesting, while the scientific potential of time domain archives (non-time-critical) is steadily increasing
- The spectroscopic follow-up crisis is going to get much worse; thus the (near)real-time classification of transients and an automated follow-up prioritization are getting even more critical
- Real-time mining of massive data streams has many applications outside astronomy