TRANSIENT SKY IN THE BIG DATA ERA

ŁUKASZ WYRZYKOWSKI (pron: Woocash Vizhikovsky)
WARSAW UNIVERSITY OBSERVATORY

with:
S. Hodgkin, M. Gromadzki, A. Hamanowicz, K. Rybicki, J. Klencki, 
Z. Kostrzewa-Rutkowska, OGLE and Gaia Teams.

EWASS 2017, Prague
S14-Astroinformatics
STATS

TNS since Jan 1, 2016
Reported: 9547
classified: 879 <10%

ratio classified:
Pan-STARRS1 136/5908
GaiaAlerts 116/2224
ATLAS 143/574
ASAS-SN 190/254
STATS

TNS since Jan 1, 2016
Reported : 9547
classified : 879 <10%

ratio classified:
Pan-STARRS1 136/5908
GaiaAlerts 116/2224
ATLAS 143/574
ASAS-SN 190/254

<10% classified in total*
75% bright classified

* not all classifications reported to TNS yet
GO BRIGHT?

bright is good:

- easy to detect
- easy to follow-up for long time until late epochs
- studies of detailed aspects of explosions
GO BRIGHT?

bright is good:

- easy to detect
- easy to follow-up for long time until late epochs
- studies of detailed aspects of explosions

should we go bright only?

No!
GO BRIGHT?

bright is good:

- easy to detect
- easy to follow-up for long time until late epochs
- studies of detailed aspects of explosions

should we go bright only?

No!

GO FAINT!

- many interesting transients are faint!
- distant SN Ia - Dark Energy and cosmology
- lensed supernovae - Dark Matter distribution
- SN impostors - SN physics
- faint and fast - new rare explosions
- the unknown!
PRE-CLASSIFICATION

- **Pixel level**
- **Context level**
- **Light curve level**
- **Spectrum level**
PIXELS

easy examples

reference image  current image  difference image
PIXELS

bogus examples

reference image  current image  difference image
PIXELS
real-bogus classification

- Self-Organizing Maps on difference images
- single image used only - allows rapid detections of transients!
- 3% false positive rate
• wealth of archives with imaging and catalogues
• problem: missing data (e.g., SDSS imaging only in the North)
• Decision Trees, Fuzzy SVM
• wealth of archives with imaging and catalogues
• problem: missing data (e.g., SDSS imaging only in the North)
• Decision Trees, Fuzzy SVM

multi-wavelength surveys:
Rosat (X)
GALEX (UV)
Wise (IR)
2MASS (NIR)
large fraction of the sky!

Assef+2013
CONTEXT

- wealth of archives with imaging and catalogues
- problem: missing data (e.g., SDSS imaging only in the North)
- Decision Trees, Fuzzy SVM

multi-wavelength surveys:
Rosat (X)
GALEX (UV)
Wise (IR)
2MASS (NIR)
large fraction of the sky!

large flare missed in a known AGN z=0.324

Assef+2013
VERON_AGN (<3'')

GSC_STAR (<3'')  USNO_STAR (<3'')  SDSS_STAR (<3'')

GCVS_VARSTAR (<3'')

LEDA_GALAXY (<3'')  SDSS_GALAXY (<3'')  2MASS_GALAXY (<3'')

Y  N

LEDA_GALAXY (3''-60'')  SDSS_GALAXY (3''-15'')  2MASS_GALAXY (3''-15'')

Decision

good match to known AGN?

star very nearby?

known variable star?

very close to galaxy core?

galaxy nearby?
• simple parameters:
  - last non-detection
  - rising slope
  - max magnitude
• training set: 90 spectroscopically classified supernovae
• classifier: Random Forest
• correct answers: 84%

\[ \text{SN Ia} \quad \text{Novae} \quad \text{SN II} \quad \text{DN} \quad \text{LPVar} \]
test on Stripe 82

Classifier does not know about the galaxy!
new supernovae in Stripe 82

overlooked super-luminous supernova found: Z.Kostrzewa-Rutkowska+2013
5 million objects per frame and ~1 star is brighter due to microlensing
Supervised classification on 150 million light curves
Random Forest with 26 attributes

3500 standard events
Hunt for black hole lenses - very long events
Pre-filtering for transients: skewness-vonNeumann
Random Forest to pick asymmetric light curves

99.9% BH candidate
M=6.9 M_\odot, D=1.9 kpc

nearby WD/NS/MS candidate
M= \sim 1.3 M_\odot, D=\sim 0.6 kpc
SPECTRA (LOW RESOLUTION)

Gaia delivers low-res (R<100) spectra for ALL OBSERVATIONS powerful to recognise SN Ia, SN II, CV, stars training on thousands of known transients spectra in high-res.

Confusion Matrix for G=19mag.
46.95% unclassified (BB or Ambig)
89.21% of true positives in classified

Blagorodnova+2014
also SED machine, Blagorodnova+2017 in prep.
SUMMARY

• to exploit science with transients we NEED Machine Learning NOW!

• individual solutions and ideas exist:
  - pixel level
  - context level (archives)
  - light curve level
  - spectra level

• is there a need for a clever, self-adjusting, unified and ultimate classification tool for generic application for transients?