Search for UV Ceti type stars in astronomical surveys

Using machine learning methods with Python

Jan Okleštěk[1], Viktor Votruba[1]

[1]Department of Theoretical Physics and Astrophysics
Masaryk University, Brno

30. June 2016

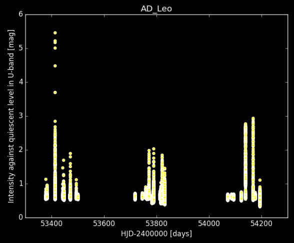
Overview

1. Flare stars
2. Methods

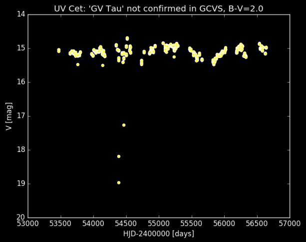
Overview

Big family of stars

- stars along main sequence
- from WR to M types
- unpredictable increases in brightness for a few minutes
- Flares detectable from X-ray to radiowaves
- largest numbers of flare stars among UV Ceti



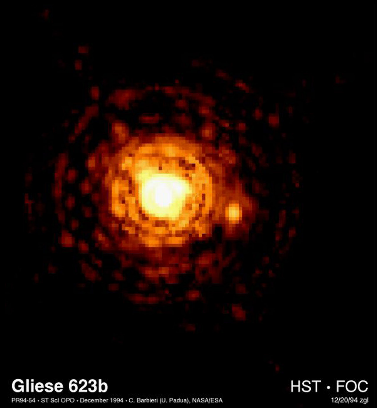
Flare star lightcurve (UV Ceti)



Misclassified UV Ceti

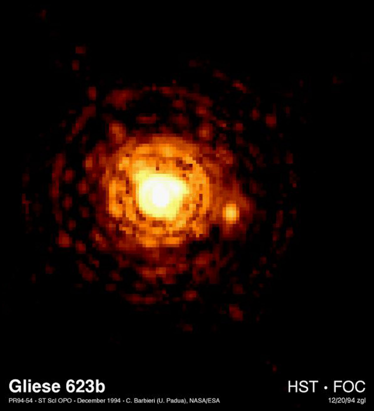
UV Ceti

- Main sequence M type stars



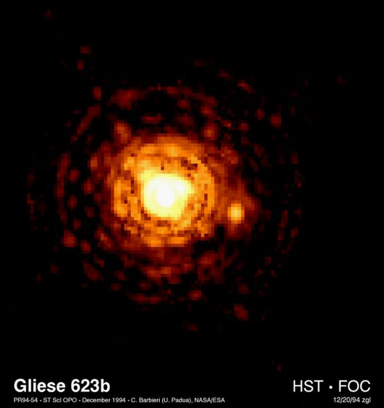
UV Ceti

- Main sequence M type stars
- low mass, slow evolution, convection



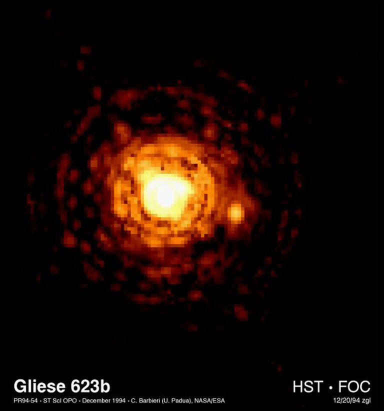
UV Ceti

- Main sequence M type stars
- low mass, slow evolution, convection
- $0.1 - 1.0\ M_\odot$, $10^{10}\ M_\odot$ (Mirzoyan, 1995)



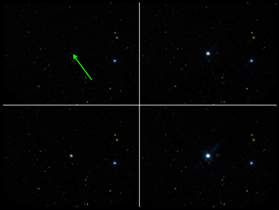
UV Ceti

- Main sequence M type stars
- low mass, slow evolution, convection
- $0.1 - 1.0\ M_{\odot}$, $10^{10}\ M_{\odot}$ (Mirzoyan, 1995)
- Luyten 726-8 (1948) - prototype

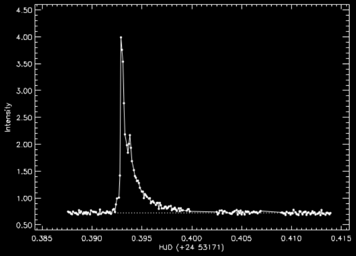


Where and how to find UV Ceti

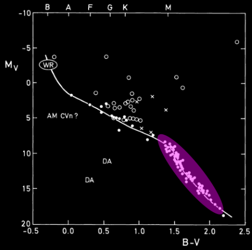
- mostly in binary systemes
- AD Leo, EV Lac, EQ Peg, V1054 Oph (Dal and Evren, 2012)
- Slow flares(<1400 s, 1 mag), Fast (<400 s, 4 mag), or 3.5 ratio of rise against decays



GJ 3685A,Photo from NASA Jet Propulsion Laboratory



Fast flare in U band, V1054 Oph (Dal and Evren, 2012)

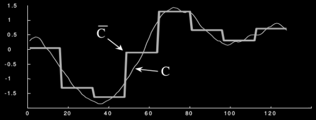


Petterson, 1989

Overview

SAX - Symbolic Aggregate approXimation

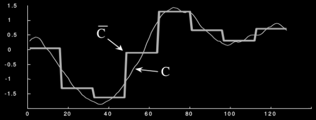
- too many points to measure in time series
- use PAA to reduce data $n \rightarrow w$



Piecewise Aggregate approximation

SAX - Symbolic Aggregate approXimation

- too many points to measure in time series
- use PAA to reduce data $n \rightarrow w$



Piecewise Aggregate approximation

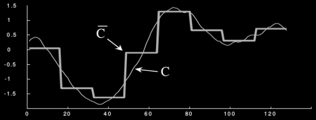
- get the distribution
- divide it equiprobably
- for different lengths of alphabet

β_i \ a	3	4	5	6	7	8	9	10
β_1	-0.43	-0.67	-0.84	-0.97	-1.07	-1.15	-1.22	-1.28
β_2	0.43	0	-0.25	-0.43	-0.57	-0.67	-0.76	-0.84
β_3		0.67	0.25	0	-0.18	-0.32	-0.43	-0.52
β_4			0.84	0.43	0.18	0	-0.14	-0.25
β_5				0.97	0.57	0.32	0.14	0
β_6					1.07	0.67	0.43	0.25
β_7						1.15	0.76	0.52
β_8							1.22	0.84
β_9								1.28

Breakpoints table

SAX - Symbolic Aggregate approXimation

- too many points to measure in time series
- use PAA to reduce data $n \to w$



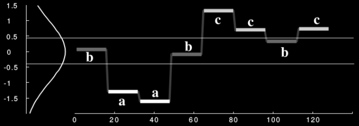
Piecewise Aggregate approximation

- get the distribution
- divide it equiprobably
- for different lengths of alphabet

β_i \ a	3	4	5	6	7	8	9	10
β_1	-0.43	-0.67	-0.84	-0.97	-1.07	-1.15	-1.22	-1.28
β_2	0.43	0	-0.25	-0.43	-0.57	-0.67	-0.76	-0.84
β_3		0.67	0.25	0	-0.18	-0.32	-0.43	-0.52
β_4			0.84	0.43	0.18	0	-0.14	-0.25
β_5				0.97	0.57	0.32	0.14	0
β_6					1.07	0.67	0.43	0.25
β_7						1.15	0.76	0.52
β_8							1.22	0.84
β_9								1.28

Breakpoints table

- assign letters to values between breakpoints
- convert time series to strings



SAX - assigning letters (J. Lin, E. Keogh, S. Lonardi, and B. Chiu, 2003)

SAX - Distance measures

- Euclidean distance (A):

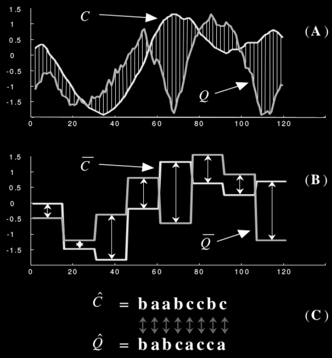
$$D(Q,C) \equiv \sqrt{\sum_{i=1}^{n}(q_i - c_i)^2}$$

- Piecewise Aggregate Approximation distance (B):

$$D(\bar{Q},\bar{C}) \equiv \sqrt{\frac{n}{w}} \cdot \sqrt{\sum_{i=1}^{w}(\bar{q}_i - \bar{c}_i)^2}$$

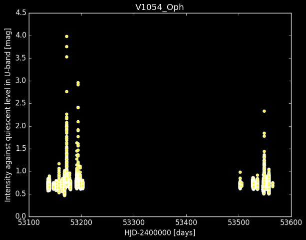
- Distance in symbolic space (C):

$$D(\hat{Q},\hat{C}) \equiv \sqrt{\frac{n}{w}} \cdot \sqrt{\sum_{i=1}^{w}(\mathrm{dist.}(\bar{q}_i, \bar{c}_i))^2}$$

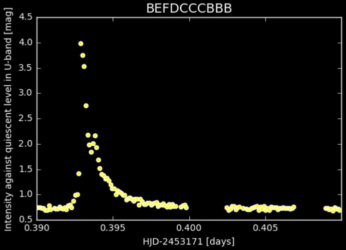


Euclidean distance, PAA and SAX (J. Lin, E. Keogh, S. Lonardi, and B. Chiu, 2003)

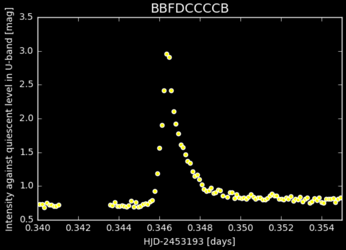
SAX - illustration



The flare star lightcurve



The biggest flare in this data (Vo, 2016)



The 2nd biggest flare in this data (Vo, 2016)

Status and future plans

Status:

- searched GCVS, VSX and other catalogs for UV Ceti training data with astropy
- found 429 stars with lightcurves from Catalina Sky Survey (Drake, A.J. et al., 2009) and their color indices in other catalogs
- found more than 2400 lightcurves from OGLE catalog for testing data

Future plans:

- find better training data for searching for UV Ceti in Kepler catalog
- get more testing data from more catalogs
- use that data to create a filter with already existing methods in LightCurvesClassifier (Vo, 2016)
- create own methods and filters for LightCurvesClassifier

Thanks!

Thank you for your attention!

References

Mirzoyan, L. (1995). The Red Dwarf Star Population in the Galaxy. International Astronomical Union Colloquium, 151, 55-56. doi:10.1017/S0252921100034564

Dal, H. A. and Evren, S. (2012) A New Method for Classifying Flares of UV Ceti Type Stars: Differences Between Slow and Fast Flares, arXiv:1206.5791v1 [astro-ph.SR]

Petterson, B.R., "A review of stellar flares and their characteristics" in Solar Physics vol. 121 pp 299, 1989

J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery, pp. 2–11, ACM, 2003

Martin Vo, "LightCurvesClassifier", at http://vocloud-dev.asu.cas.cz/lcc/ and https://github.com/mavrix93/LightCurvesClassifier, 2016

Drake, A.J. et al. First Results from the Catalina Real-time Transient Survey 2009, ApJ, 696, 870