

ИЗВЛЕЧЕНИЕ ПРИЗНАКОВ ИЗ КРИВЫХ БЛЕСКА АСТРОНОМИЧЕСКИХ ИСТОЧНИКОВ

А. Д. Лаврухина¹, К. Л. Маланчев^{2,3}

¹*Московский государственный университет им. М. В. Ломоносова, факультет космических исследований,* ²*Московский государственный университет им. М. В. Ломоносова, Государственный астрономический институт им. П. К. Штернберга,* ³*Department of Astronomy, University of Illinois at Urbana-Champaign*

Астрономия входит в эпоху больших обзоров переменного неба, таких как Zwicky Transient Facility и грядущий Legacy Survey of Space and Time. Эти обзоры предназначены для мониторинга всей видимой части неба за несколько ночей. Для эффективной работы с такими объемами данных требуется автоматическое извлечение признаков кривых блеска для дальнейшего их использования в задачах классификации и определения физических параметров переменных источников. В данной работе мы представляем новую библиотеку на языке Python, предназначенную для извлечения 25 признаков кривых блеска переменных астрономических источников.

LIGHT CURVE FEATURES EXTRACTION FROM ASTRONOMICAL SOURCE

A. D. Lavrukhina¹, K. L. Malanchev^{2,3}

¹*Lomonosov Moscow State University, Faculty of Space Research,* ²*Lomonosov Moscow State University, Sternberg astronomical institute,* ³*Department of Astronomy, University of Illinois at Urbana-Champaign*

Astronomy is entering the era of large surveys of the variable sky such as Zwicky Transient Facility and forthcoming Legacy Survey of Space and Time. These surveys are dedicated to observe the apparent celestial sphere in several days. The usage of an automatic extraction of light curve features is required to work effectively with such amount of data, especially in problems of classification and characterisation of variable sources. In this work, we present the new Python library for extraction of 25 light curve features of variable astronomic sources.

Современные астрономические обзоры содержат информацию о сотнях миллионов кривых блеска переменных астрономических источников, например, релиз данных Zwicky Transient Facility Data Release 3 (ZTF DR3) содержит миллиарды кривых блеска. При решении задач классификации [1] или поиска аномалий [2] в таких больших объемах данных используются методы машинного обучения. Обычно кривые блеска не используются напрямую, вместо этого каждый источник представляется набором признаков, которые наилучшим образом описывают свойства его переменности. В данной работе мы представляем новую библиотеку на языке Python, предназначенную для извлечения признаков из кривых блеска переменных астрономических источников.

Пусть кривая блеска задана набором из N наблюдений $\{t_i, m_i, \delta_i\}$, где t_i — время наблюдения; m_i — наблюдаемая звездная величина; δ_i — ошибка наблюдения звездной величины.

Нами рассмотрены следующие признаки (мы придерживаемся терминологии оригинальных статей и других библиотек, таких как Feets [3]):

- Amplitude — полуамплитуда кривой блеска.

- Anderson–Darling test — статистика теста Андерсона–Дарлинга для нормального распределения.
- Beyond n std — доля наблюдений, отклонившихся от среднего значения звездной величины более чем на n среднеквадратических отклонений [4].

- Cusum — размах частичных сумм вида $S_j \equiv \frac{1}{N\sigma_m} \sum_{i=0}^j (m_i - \langle m \rangle)$ [1].

- Von Neuman η [5]

$$\eta \equiv \frac{1}{(N-1)\sigma_m^2} \sum_{i=0}^{N-2} (m_{i+1} - m_i)^2.$$

- η^e — обобщение коэффициента η для неравномерных временных рядов [1]

$$\eta^e \equiv \frac{(t_{N-1} - t_0)^2}{(N-1)^3} \frac{\sum_{i=0}^{N-2} \left(\frac{m_{i+1} - m_i}{t_{i+1} - t_i} \right)^2}{\sigma_m^2}.$$

- Excess variance — коэффициент изменчивости амплитуды звездной величины [6]

$$\frac{\sigma_m^2 - \langle \delta^2 \rangle}{\langle m \rangle^2},$$

где $\langle \delta^2 \rangle$ — среднее квадрата ошибки; σ_m — среднеквадратическое отклонение звездной величины; $\langle m \rangle$ — среднее значение звездной величины.

- Inter-percentile range — интерпроцентильный размах (частный случай — интерквартильный размах).
- Kurtosis — коэффициент эксцесса.
- Linear fit — линейная аппроксимация данных с учетом известных ошибок наблюдений.
- Linear trend — линейная аппроксимация данных в предположении о шуме с неизвестной, но постоянной дисперсией.
- Magnitude percentage ratio [4]

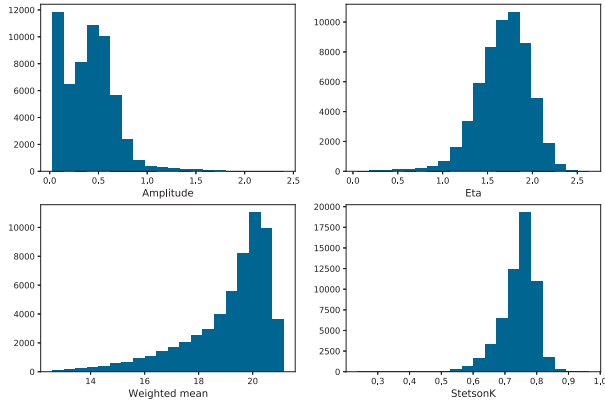
$$\frac{Q(1-n) - Q(n)}{Q(1-d) - Q(d)},$$

где $Q(n)$ и $Q(d)$ — n -й и d -й квантили выборки звездных величин.

- Maximum slope — максимальный коэффициент наклона прямой между двумя наблюдениями звездной величины

$$\max_{i=0 \dots N-2} \left| \frac{m_{i+1} - m_i}{t_{i+1} - t_i} \right|.$$

- Mean — среднее значение звездной величины.
- Mean variance — отношение среднеквадратического отклонения звездной величины к ее среднему значению.
- Median — медианное значение звездной величины.



Гистограммы распределения значений признаков для выборки кривых блеска ZTF DR3

- Median absolute deviation — медианное значение отклонений звездной величины от ее медианы [4]

$$\text{Median}(|m_i - \text{Median}(m)|).$$

- Median buffer range percentage — доля наблюдений, попавшая в интервал $\text{Median}(m) \pm q \times (\max(m) - \min(m))/2$ [4].
- Percent amplitude — максимальное значение отклонений звездной величины от ее медианы [4].
- Percent difference magnitude percentile — отношение p -го интерперцентильного размаха к медианному значению [4]

$$\frac{Q(1-p) - Q(p)}{\text{Median}(m)}.$$

- Reduced chi-squared — редуцированная статистика критерия χ^2 Пирсона для нормального распределения.
- Skew — коэффициент асимметрии звездной величины.
- Standard deviation — среднеквадратическое отклонение звездной величины.
- Stetson K — коэффициент Стетсона K [7]

$$\frac{\sum_i \left| \frac{m_i - \bar{m}}{\delta_i} \right|}{\sqrt{N} \chi^2}.$$

- Weighted Mean — взвешенное среднее звездной величины с учетом ошибок измерений.

На рисунке показаны гистограммы распределения амплитуды, коэффициента η , взвешенного среднего и коэффициента Стетсона K для тестового набора данных из 57 546 кривых блеска ZTF DR3 в фильтре zr из поля 695.

В дальнейшем мы планируем реализовать функции для извлечения других признаков, таких как период кривой блеска, а также разработать новые признаки, подходящие для

задач классификации. Кроме того, в планах — валидация признаков при помощи решения тестовых задач классификации с помощью машинного обучения.

Репозиторий библиотеки: <https://github.com/anlava/light-curve/tree/dev>

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-02-00779, при поддержке Междисциплинарной научно-образовательной школы Московского университета «Фундаментальные и прикладные исследования космоса».

Библиографические ссылки

- [1] *Kim Dae-Won, Protopapas Pavlos, Bailer-Jones Coryn A. L. et al.* The EPOCH Project // *Astronomy & Astrophysics*. — 2014. — Vol. 566. — P. A43.
- [2] *Malanchev K. L., Pruzhinskaya M. V., Korolev V. S. et al.* Anomaly detection in the Zwicky Transient Facility DR3 // *arXiv e-prints*. — 2020. — P. arXiv:2012.01419. 2012.01419.
- [3] *Cabral J. B., Sánchez B., Ramos F. et al.* From FATS to feets: Further improvements to an astronomical feature extraction tool based on machine learning // *Astronomy and Computing*. — 2018.
- [4] *D’Isanto A., Cavuoti S., Brescia M. et al.* An analysis of feature relevance in the classification of astronomical transients with machine learning methods // *Monthly Notices of the Royal Astronomical Society*. — 2016. — Vol. 457, № 3. — P. 3119–3132. <https://academic.oup.com/mnras/article-pdf/457/3/3119/8001547/stw157.pdf>.
- [5] *Neumann John von.* Distribution of the Ratio of the Mean Square Successive Difference to the Variance // *Ann. Math. Statist.* — 1941. — Vol. 12, № 4. — P. 367–395.
- [6] *Sánchez P., Lira P., Cartier R. et al.* Near-infrared Variability of Obscured and Unobscured X-Ray-selected AGNs in the COSMOS Field // *The Astrophysical Journal*. — 2017. — Vol. 849, № 2. — P. 110.
- [7] *Stetson Peter B.* On the Automatic Determination of Light-Curve Parameters for Cepheid Variables // *Publications of the Astronomical Society of the Pacific*. — 1996. — Vol. 108. — P. 851.