



Факультет компьютерных наук НИУ ВШЭ

Scientific Open Source Meetup 2025

Санкт-Петербург
26.02.2025

Открытый код НИУ ВШЭ

Гущин Михаил, mhushchyn@hse.ru

ФКН НИУ ВШЭ



«Сухие» цифры о науке на ФКН

4 445

научных публикаций

629

публикаций Q1

168

публикаций
на конференциях A*

>85%

из списка A



Зачем нужен открытый код в науке?

- **Увеличение видимости исследований**

- Открытый код на GitHub значительно расширяет круг потенциальных пользователей вашего кода и, соответственно, читателей (исследователи и разработчики) вашей публикации.

- **Прозрачность и воспроизводимость**

- Код позволяет другим исследователям легко проверить ваши методы и результаты. Это создает больше доверия к работе.

- **Обмен знаниями и сотрудничество**

- Исследователи могут использовать уже существующие разработки, адаптируя их под свои нужды, что ускоряет научные процессы и повышает эффективность работы.

- **Экономия ресурсов**

- Чем проще повторить вашу работу, тем больше вероятность, что другие исследователи будут использовать ваши методы и ссылаться на вас в своих работах.



Цитирования и звезды

Код статей NeurIPS на GitHub

Год	Статья	Код	# цитирований на ArXiv	# форков на GitHub	# звезд на GitHub
2023	LLM-Pruner: On the Structural Pruning of Large Language Models	https://github.com/horseee/LLM-Pruner	148	74	698
	An Inverse Scaling Law for CLIP Training	https://github.com/UCSC-VLAA/CLIPA	19	10	280
	RayDF: Neural Ray-surface Distance Fields with Multi-view Consistency	https://github.com/vLAR-group/RayDF	2	4	102
2022	VideoMAE: Masked Autoencoders are Data-Efficient Learners for Self-Supervised Video Pre-Training	https://github.com/MCG-NJU/VideoMAE	648	123	1241
	Gradient Descent: The Ultimate Optimizer	https://github.com/kach/gradient-descent-the-ultimate-optimizer	34	25	358
	GenerSpeech: Towards Style Transfer for Generalizable Out-Of-Domain Text-to-Speech	https://github.com/Rongjiehuang/GenerSpeech	62	44	307
2021	Rethinking Space-Time Networks with Improved Memory Coverage for Efficient Video Object Segmentation	https://github.com/hkchengrex/STCN	272	71	525
	Prototypical Cross-Attention Networks for Multiple Object Tracking and Segmentation	https://github.com/SysCV/pcan	82	50	360
	Deceive D: Adaptive Pseudo Augmentation for GAN Training with Limited Data	https://github.com/EndlessSora/DeceiveD	99	24	251



Мировые примеры: Ruptures

Signal Processing 167 (2020) 107299



Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro



Review

Selective review of offline change point detection methods

Charles Truong^a, Laurent Oudre^{b,*}, Nicolas Vayatis^a

^aCMLA, CNRS, ENS Paris Saclay France
^bL2TI, University Paris 13 France



ARTICLE INFO

Article history:
Received 14 December 2018
Revised 19 August 2019
Accepted 16 September 2019
Available online 19 September 2019

Keywords:
Change point detection
Segmentation
Statistical signal processing

ABSTRACT

This article presents a selective survey of algorithms for the offline detection of multiple change points in multivariate time series. A general yet structuring methodological strategy is adopted to organize this vast body of work. More precisely, detection algorithms considered in this review are characterized by three elements: a cost function, a search method and a constraint on the number of changes. Each of those elements is described, reviewed and discussed separately. Implementations of the main algorithms described in this article are provided within a Python package called ruptures.

© 2019 Elsevier B.V. All rights reserved.

- C. Truong, L. Oudre, N. Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020
 - 1360 цитирований (данные Google Scholar)

ruptures Search deepcharles/ruptures v1.1.9 1.7k 162

Home Getting started User guide Gallery of examples Code reference About

Home
Welcome to ruptures

Welcome to ruptures

Maintained? **yes** build **falling** python 3.6 | 3.7 | 3.8 | 3.9 pypi package 1.1.9 conda-forge v1.1.9 docs **passing**
 license BSD-2-Clause downloads 20M code style black launch binder Codecov

ruptures is a Python library for off-line change point detection. This package provides methods for the analysis and segmentation of non-stationary signals. Implemented algorithms include exact and approximate detection for various parametric and non-parametric models. ruptures focuses on ease of use by providing a well-documented and consistent interface. In addition, thanks to its modular structure, different algorithms and models can be connected and extended within this package.

How to cite. If you use ruptures in a scientific publication, we would appreciate citations to the following paper:

- C. Truong, L. Oudre, N. Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020. [journal] [pdf]

- Table of contents
- Latest news
- Basic usage
- General information
- Contact
- Important links
- Dependencies and install
- Changelog
- Thanks to all our contributors
- License

- Библиотека алгоритмов из статьи: <https://github.com/deepcharles/ruptures>
 - 1700+ звезд на GitHub
 - 20M скачиваний с PyPi



Мировые примеры: MIT HAN LAB

- <https://github.com/mit-han-lab>
- Официальный репозиторий лаборатории Hardware Accelerated Neural-nets в MIT
- Выкладывают код своих исследований и библиотеки алгоритмов
- Результаты исследований презентуют на Core A/A* конференциях: NeurIPS, ICML, CVPR, ICLR, ICCV и других
- Репозитории набирают больше звезд, чем цитирований у статей (это нормально)

mit-han-lab

Overview Repositories 58 Projects Packages People 1

MIT HAN Lab
Efficient AI Computing. PI: Song Han
3k followers MIT <https://hanlab.mit.edu> @songhan_mit

Unfollow

Pinned

- streaming-llm** (Public)
[ICLR 2024] Efficient Streaming Language Models with Attention Sinks
Python 6.8k 377
- llm-awq** (Public)
[MLSys 2024 Best Paper Award] AWQ: Activation-aware Weight Quantization for LLM Compression and Acceleration
Python 2.8k 232
- efficientvit** (Public)
Efficient vision foundation models for high-resolution generation and perception.
Python 2.7k 213
- bevfusion** (Public archive)
[ICRA'23] BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation
Python 2.5k 444
- temporal-shift-module** (Public)
[ICCV 2019] TSM: Temporal Shift Module for Efficient Video Understanding
Python 2.1k 420
- once-for-all** (Public)
[ICLR 2020] Once for All: Train One Network and Specialize it for Efficient Deployment
Python 1.9k 338

People

Top languages

- Python C++ Jupyter Notebook
- Cuda C

Most used topics

- pytorch deep-learning efficient-model
- acceleration quantization

Report abuse



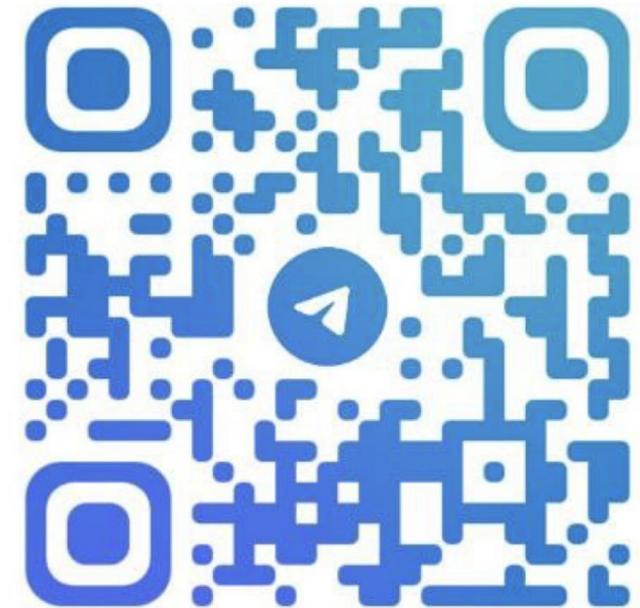
Открытый код ФКН

Цели проекта по открытому коду ФКН:

- Увеличить видимость и узнаваемость ФКН на GitHub
- Развить сообщество по открытому коду внутри ФКН
- Повысить интеграцию ФКН в мировое сообщество по открытому коду

Что сделали и делаем:

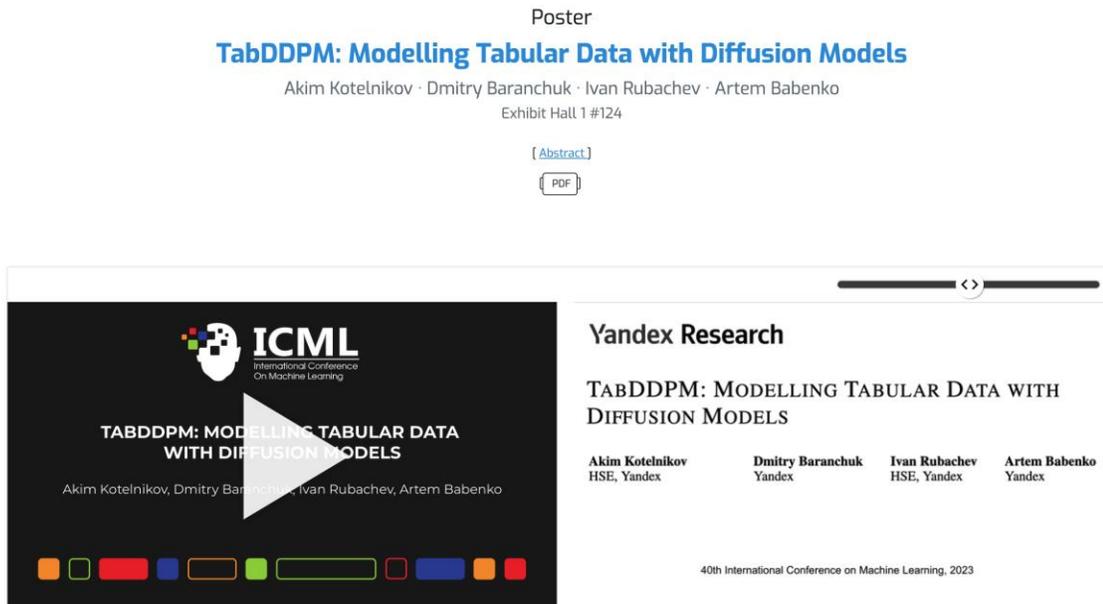
- Создали каналы связи: канал ТГ, GitHub, сайт, почта
- Рассказали о 40+ проектах сотрудников и студентов ФКН
- Информлируем о конкурсах и грантах
- Анонсы мероприятий



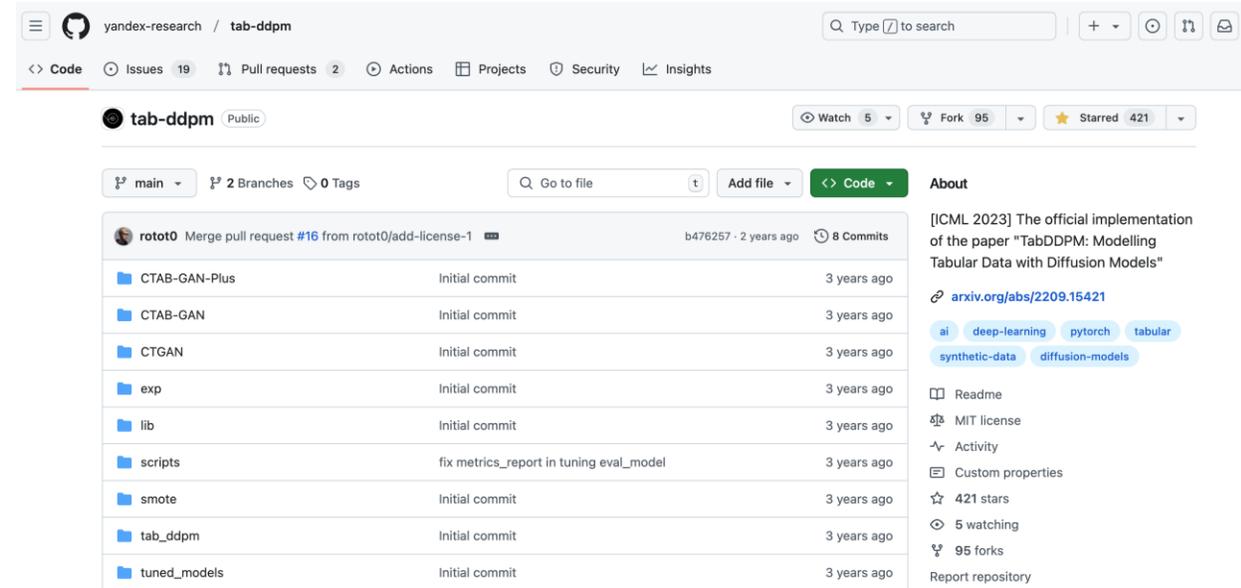
@HSE_CS_OPENSOURCE



Наши примеры: TabDDPM



- Лаборатория компании Яндекс на ФКН НИУ ВШЭ
- Текст и доклад на ICML: <https://icml.cc/virtual/2023/poster/24703>
 - 270 цитирований (данные Google Scholar)



- Код алгоритмов и экспериментов из статьи: <https://github.com/yandex-research/tab-ddpm>
 - 400+ звезд на GitHub
 - 90+ форков репозитория
- Одна из наиболее популярных моделей в мире



Наши примеры: Roerich

Journal of Computational Science 53 (2021) 101385



ELSEVIER

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs



Generalization of change-point detection in time series data based on direct density ratio estimation

Mikhail Hushchyn*, Andrey Ustyuzhanin

HSE University, 20 Myasnitskaya Ulitsa, Moscow 101000, Russia

ARTICLE INFO

Keywords:
Time series
Change-point detection
Machine learning
Neural networks
Density ratio

ABSTRACT

The goal of the change-point detection is to discover changes of time series distribution. One of the art approaches of change-point detection is based on direct density ratio estimation. In this work, we show how existing algorithms can be generalized using various binary classification and regression models. In particular, we show that the Gradient Boosting over Decision Trees and Neural Networks can be used for this purpose. The algorithms are tested on several synthetic and real-world datasets. The results show that the proposed methods outperform classical RuLSIF algorithm. Discussion of cases where the proposed algorithms have advantages over existing methods is also provided.

- Лаборатория методов анализа больших данных (Lambda) ФКН
- Текст: <https://arxiv.org/abs/2001.06386>
 - 35 + 19 цитирований (данные Google Scholar)

The screenshot shows the Roerich library website. At the top, there is a search bar and a 'Roerich' logo. Below the search bar, there is a 'Home' button and a list of links: 'Dependencies and install', 'Basic usage', 'Support', 'Related libraries', and 'Thanks to all our contributors'. A 'Check for updates' button is also visible. The main content area lists 'Change point' algorithms: 'DETECTION ALGORITHMS', 'ChangePointDetectionClassifier', 'ChangePointDetectionClassifierCV', 'OnlineNNClassifier', 'SlidingWindows', 'EnergyDistanceCalculator', 'ChangePointDetectionRuLSIF', and 'OnlineNNRuLSIF'. Below this, there is an 'EXPLANATION' section with 'MatrixImportance' and a 'DENSITY RATIO' section with 'GBDTRuLSIFRegressor' and 'NNRuLSIFRegressor'.

The screenshot shows the GitHub page for the Roerich library. At the top, there is a 'Home' button and an 'Edit on GitHub' button. The main heading is 'Welcome to Roerich'. Below this, there are statistics: 'pypi package 0.6.0', 'documentation yes', 'downloads 12k', and 'License BSD 2-Clause'. The text describes the library as a python library for online and offline change point detection for time series analysis, signal processing, and segmentation. It was named after the painter Nicholas Roerich, known as the Master of the Mountains. A link to the website is provided: <https://www.roerich.org>. Below the text is a large image of a mountain range, captioned 'Fragment of "Himalayas", 1933'. At the bottom, there is a list of articles: 'Mikhail Hushchyn and Andrey Ustyuzhanin. "Generalization of Change-Point Detection in Time Series Data Based on Direct Density Ratio Estimation." J. Comput. Sci. 53 (2021): 101385. [journal] [arxiv]' and 'Mikhail Hushchyn, Kenenbek Arzymbatov and Denis Derkach. "Online Neural Networks for Change-Point Detection." ArXiv abs/2010.01388 (2020). [arxiv]'.

- Код алгоритмов из статей: <https://github.com/HSE-LAMBDA/roerich>
 - 60+ звезд на GitHub

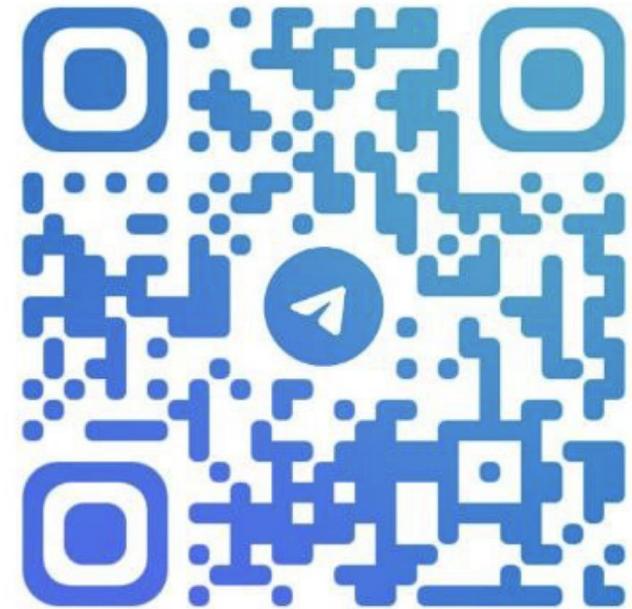


С чем мы столкнулись и еще не нашли решений

- Как мотивировать сотрудников выкладывать код?
 - Публикация статей учитывается в KPI сотрудников. Но формальных требований к коду нет.
- Как мотивировать студентов выкладывать свой код?
 - Нужно действовать только мягкой силой? Или обязать студентов выкладывать код своих КР и ВКР?
- Где выкладывать открытый код?
 - Чаще код проектов выкладывают в личных аккаунтах и на GitHub лабораторий/отделов. Что нужно выкладывать на GitHub факультета?



Подпишитесь на наш канал 😊



@HSE_CS_OPENSOURCE





Организации на GitHub (# звезд самого популярного проекта)

Зарубежные университеты:

- <https://github.com/stanfordnlp> (9343*)
- <https://github.com/mit-han-lab> (5894*)
- <https://github.com/thu-coai> (1639*)
- <https://github.com/AllenInstitute> (316*)

Зарубежные организации:

- <https://github.com/openai> (56k*)
- <https://github.com/google-research> (32k*)
- <https://github.com/facebook> (219k*)
- <https://github.com/apple> (65k*)

Организации РФ:

- <https://github.com/yandexdataschool> (9270)
- <https://github.com/Yandex> (8110*)
- <https://github.com/yandex-research> (942*)
- <https://github.com/AIRI-Institute> (187*)
- <https://github.com/sb-ai-lab> (1307*)
- <https://github.com/MobileTeleSystems> (311*)
- <https://github.com/tinkoff-ai> (1140*)

Университеты РФ:

- <https://github.com/SciProgCentre> (672*)
- <https://github.com/aimclub> (659*)
- <https://github.com/SkoltechRobotics> (187*)

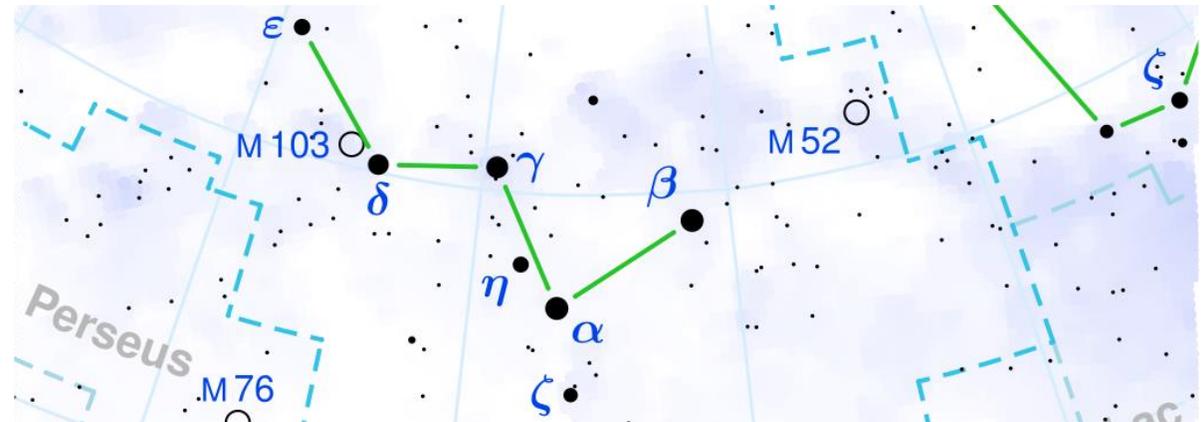
НИУ ВШЭ:

- <https://github.com/bayesgroup> (1046*)
- <https://github.com/HSE-LAMBDA> (111*)
- <https://github.com/Genomics-HSE> (15*)
- <https://github.com/robotics-laboratory> (14*)



Fulu

- <https://github.com/hse-cs/fulu>
- Библиотека методов аппроксимации кривых блеска астрономических объектов с использованием нейронных сетей.
- Названа в честь звезды Дзета Кассиопеи в 590 световых годах от нас, которая имеет официальное название Fulu
- Разработана в результате совместного научного исследования 7 организаций из 3 стран



LaNeta

- <https://github.com/hse-cs/LaNeta>
- Библиотека для оценки времен примешивания между двумя популяциями при двух пульсах миграции.
- Позволяет точно исследовать недавнюю (в пределах нескольких десятков поколений) историю примешивания популяций в сложных сценариях, для которых существовавшие ранее методы были неприменимы или неточны.

