Advanced Time Series

Lecture 1: **Introduction to time series**

Gleb Ivashkevich

whoami

Gleb Ivashkevich

doing **deep learning** - time series, satellite imagery

PhD in theoretical physics

6 years in **academia** doing numerical simulations

7 years in **data science** and **machine learning**

datarythmics **effimly**

data driven manufacturing efficiency

Our TA

Anatoly Bardukov

doing image similarity search in Yandex

applied math from HSE

back/front-end, data engineering

rocket launch: 6 тыс изображе

yandex.ru/images/search?text=rocket%20launch&from=tabbar&p=2&pos=119&rpt=simage&img_url=https%3A%2F%2Fupload.wikimedia.org%2Fwikipedia%2Fcommons%2Fthumb%2F9%2F96%2FOrbital_Test_Vehicle_5_Mission_%252836689886...

SpaceX Aims to Launch Satellite on Used Rocket Early Monday:

ca.news.yahoo.com

SpaceX Aims to **Launch** Satellite on Used ...

Пожаловаться

X 37b ракета Фалькон

Запуск ракеты Россия

Открыть 2048×1365

Похожие

Отправить

Связанные картинки

What is ATS about?

Time series are **sequences.**

- all the sequential **deep learning** blocks can be applied (RNN, CNN, combinations, etc.)
- **various problems:** forecasting, classification, event prediction
- **mental image** of time series problems - how to select parameters, architectures, etc.

What ATS is not about?

- AR, VAR, ARIMA, etc.
- specific domains
- econometrics
- financial time series

But we will **overview** the classical stuff. And will use it to build **intuition**.

You should attend if...

... most of the following is true:

- you have some basic experience with classical time series approaches
- know how to read research papers
- have some understanding of deep learning in general and RNN and CNN in particular
- have no problems at all with Python stack (especially PyTorch)

How do we check prerequisites

- **qualification** homework
- Pandas time series functionality + PyTorch
- **10 days to submit** (deadline is 24:00 Feb 3)
- at least **5 points out of 7**
- **individual** submissions

Course logistics

6 lectures, 3 hours each, Google classroom

- slides + notebook(-s)
- 4 homeworks (in pairs)
- 2 paper reviews + 2 paper implementations
- manual grading
- Slack

Course structure

Intro + 4 topics:

- TS **forecasting**: RNNs$^{\text{(various)}}$
- TS **classification**: CNN$^{\text{(various)}}$ and combined$^{\text{(RNN+CNN)}}$ models
- **TTE prediction**: DL models for predictive maintenance, survival analysis
- TS **representation learning**: VAMPnets, autoencoders$^{\text{(various)}}$

Course structure

Power consumption
Weather
Sales
Traffic

Forecasting

AR models
RNNs
Probabilistic forecasts

ATS

Activity data (inertial sensors)
Sound
Medical signals

Classification

Dilated/causal convolutions
CNNs and hybrid models
Attention mechanism
Segmentation

Any highly structured TS

Representation

VAMPnets
AEs

Equipment data
Earthquakes and other physical signals

TTE prediction

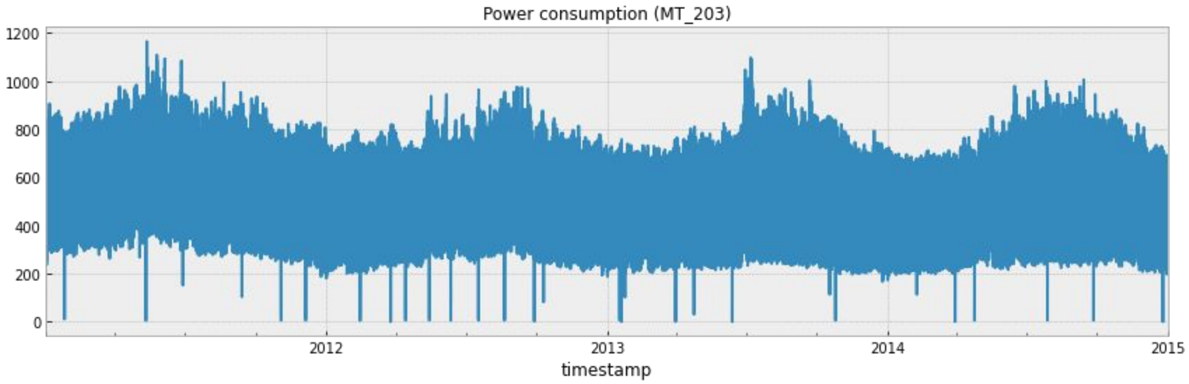
Survival models
interpretability

Time series basics: time domain

Power consumption

Electricity load [dataset]

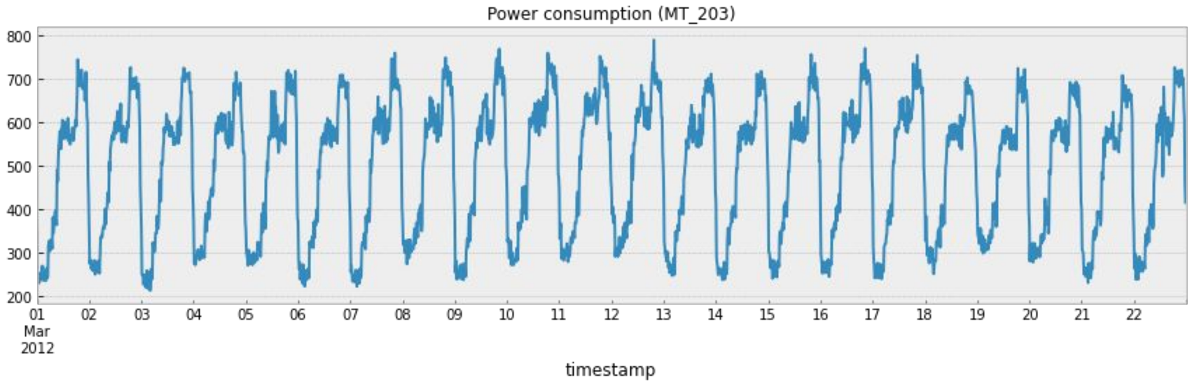
- 370 individual households
- 15 minutes sampling interval



Power consumption

Electricity load dataset

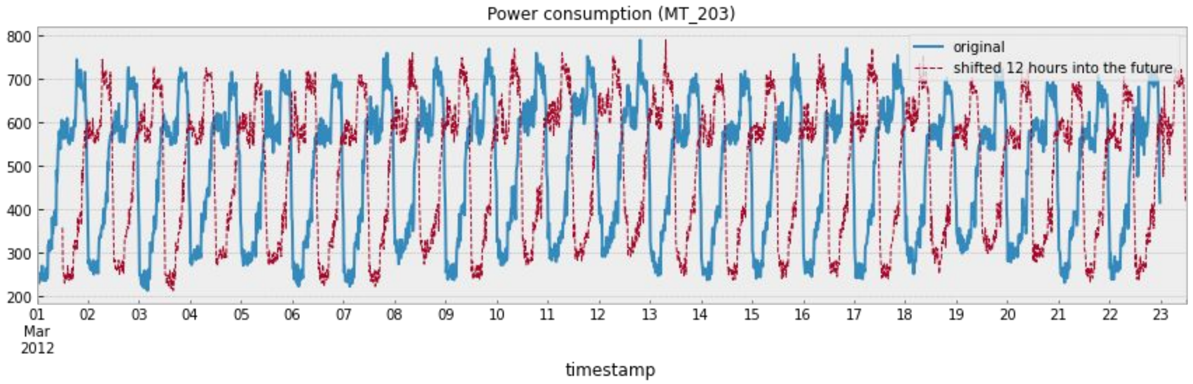
- 370 individual households
- 15 minutes sampling interval



Power consumption

Temporal structure?

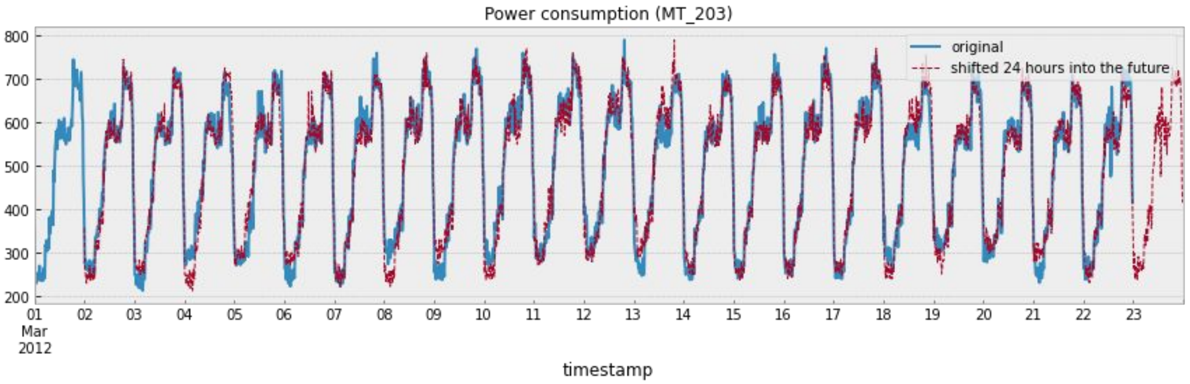
- shift the data: 12 hours



Power consumption

Temporal structure?

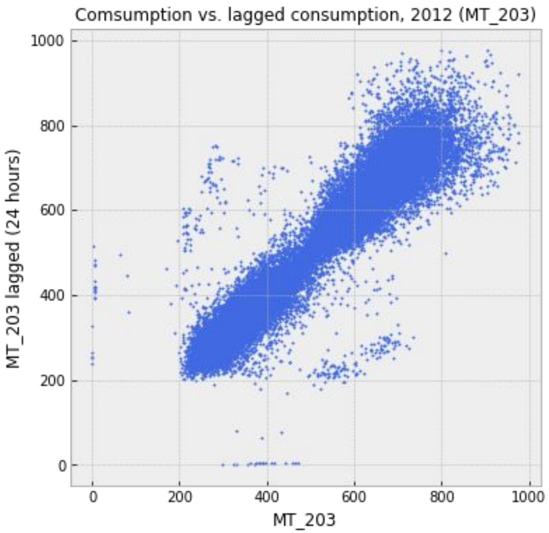
- shift the data: 24 hours

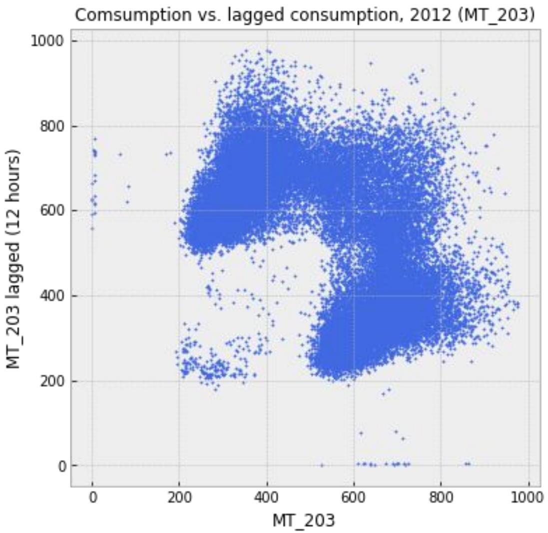


Power consumption

Temporal structure?

- shift the data: 12, 24 hours

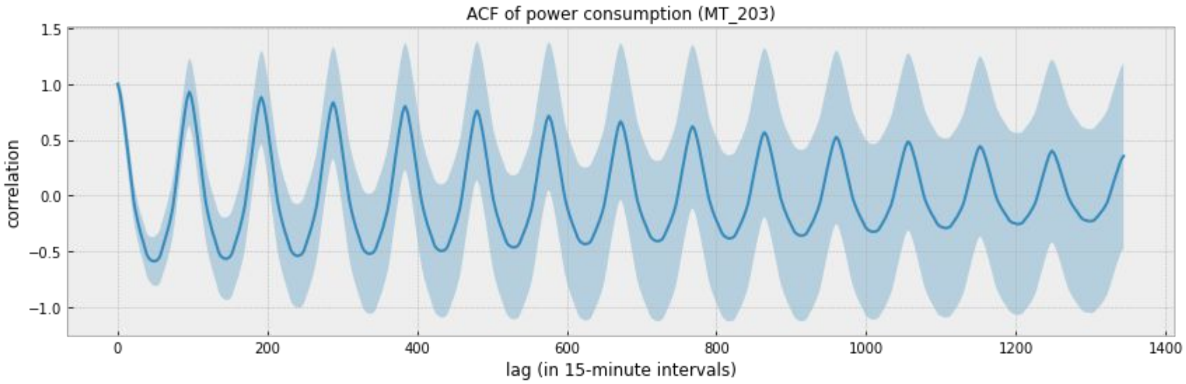




Autocorrelation function

Temporal structure?

- calculate correlations at different shifts: **autocorrelation function** (ACF)



Autocorrelation function

Temporal structure?

- calculate correlations at different shifts: **autocorrelation function** (ACF)

$$ACF(\tau) = C(y(t)y(t-\tau))$$

Autocorrelation function

ACF:

- quick assessment of signal temporal structure
- correlation length -> model parameters

$$ACF(\tau) = C(y(t)y(t-\tau))$$

Modeling

Modeling ideas:

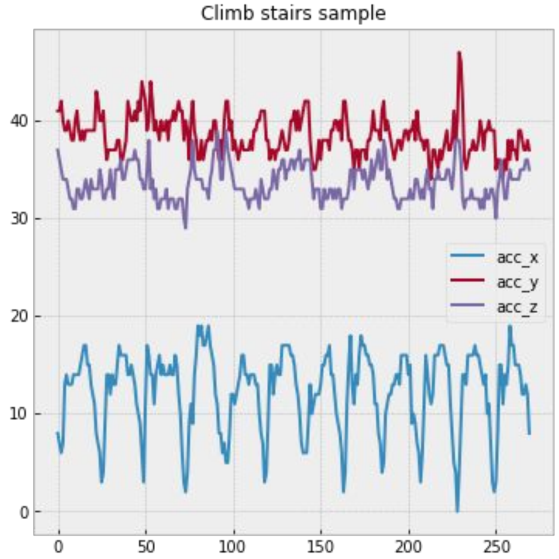
- use some AR-alike classical model
- use some RF, gradient boosting, provide lags explicitly
- **recurrent** model (modified?)

Time series basics:
frequency domain

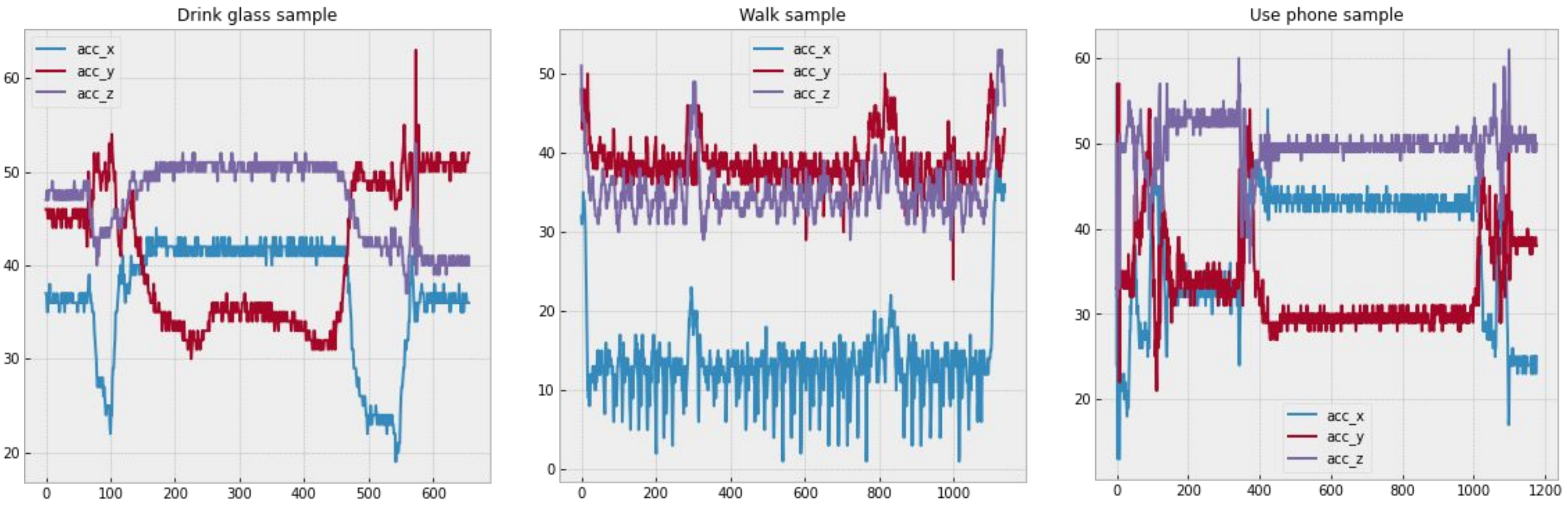
Activity recognition

ADL Recognition [dataset]

- multiple activities, short samples
- 50 Hz sampling rate



Activity recognition



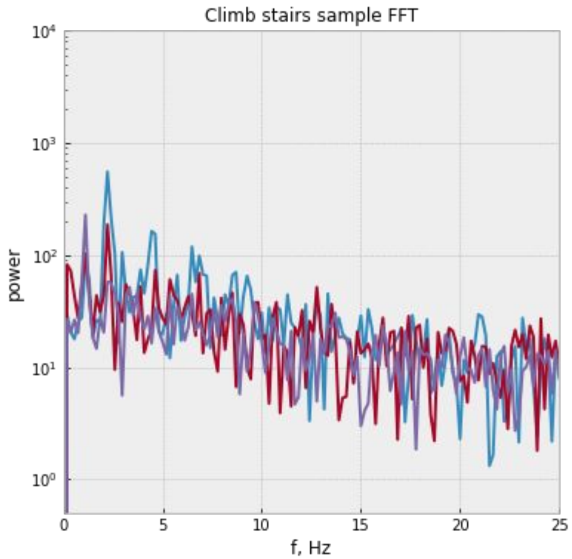
Activity recognition

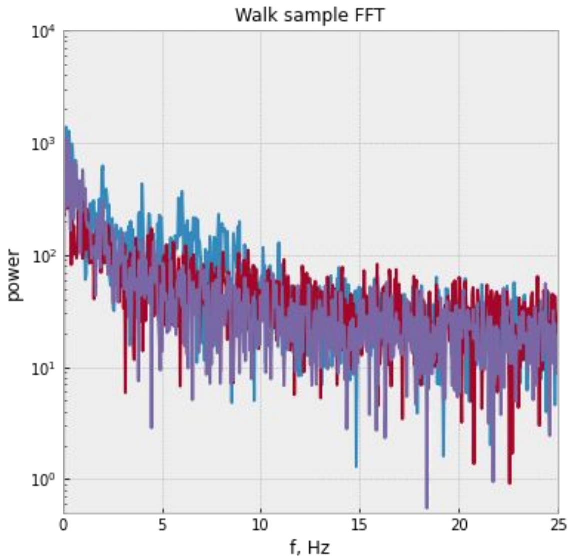
ACF? Non-informative: wrong “scale”

- FFT for the rescue

Activity recognition

Power spectrum:





Modeling

Modeling ideas:

- use manual features (`tsfresh`) and some classical model
- **CNN** model (modified?)

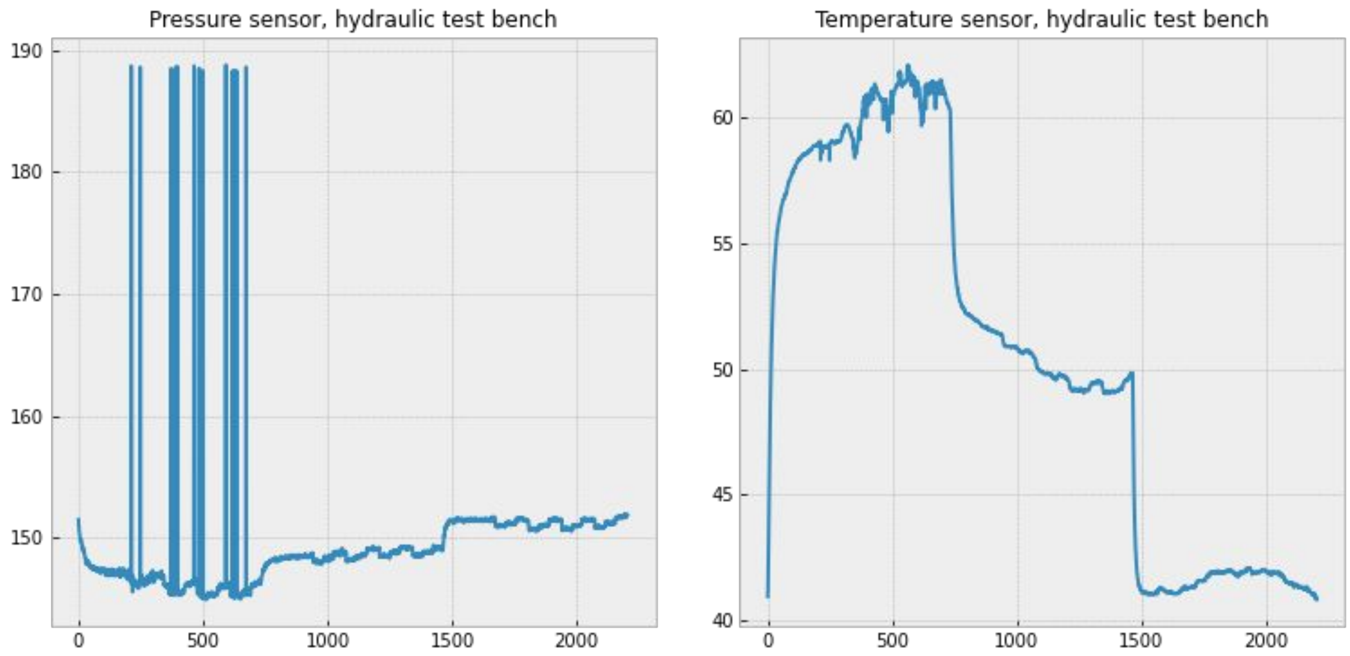
Time series basics: other examples

Industrial equipment

Condition monitoring of hydraulic systems [dataset]

- multiple sensors (temperature, pressure)
- 100 Hz sampling rate

Industrial equipment



Time series basics: tools

Pandas

Mostly for preprocessing:

- datetime operations
- resampling, rolling
- shifts

Cross-validation

Random split cannot be applied to time series

- use fixed split
- use rolling CV
- otherwise, autocorrelation will kill your model

questions?