УНИВЕРСИТЕТ ИТМО

Evaluating Time Series Predictability via Transition Graph Analysis

Anton Kovantcev, Klavdiya Bochenina, Peter Chunaev

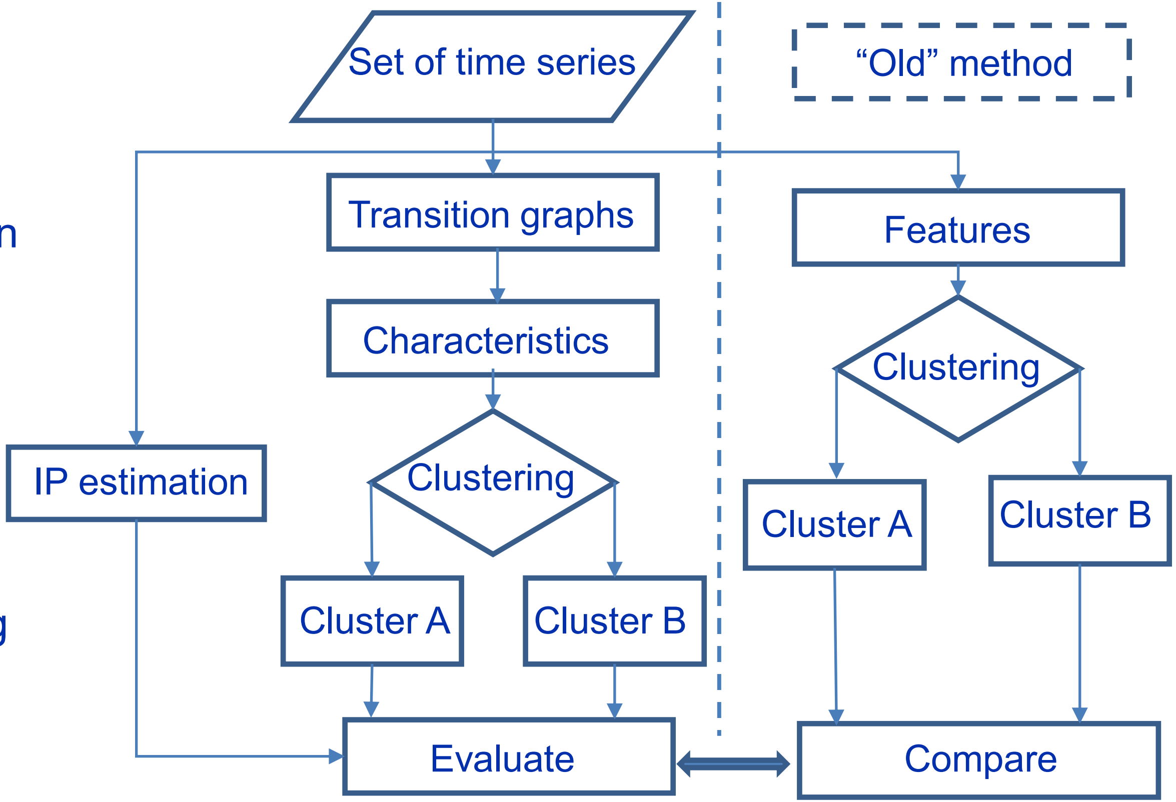
ankovantcev@itmo.ru

Introduction

Getting started with this research we want to find out:

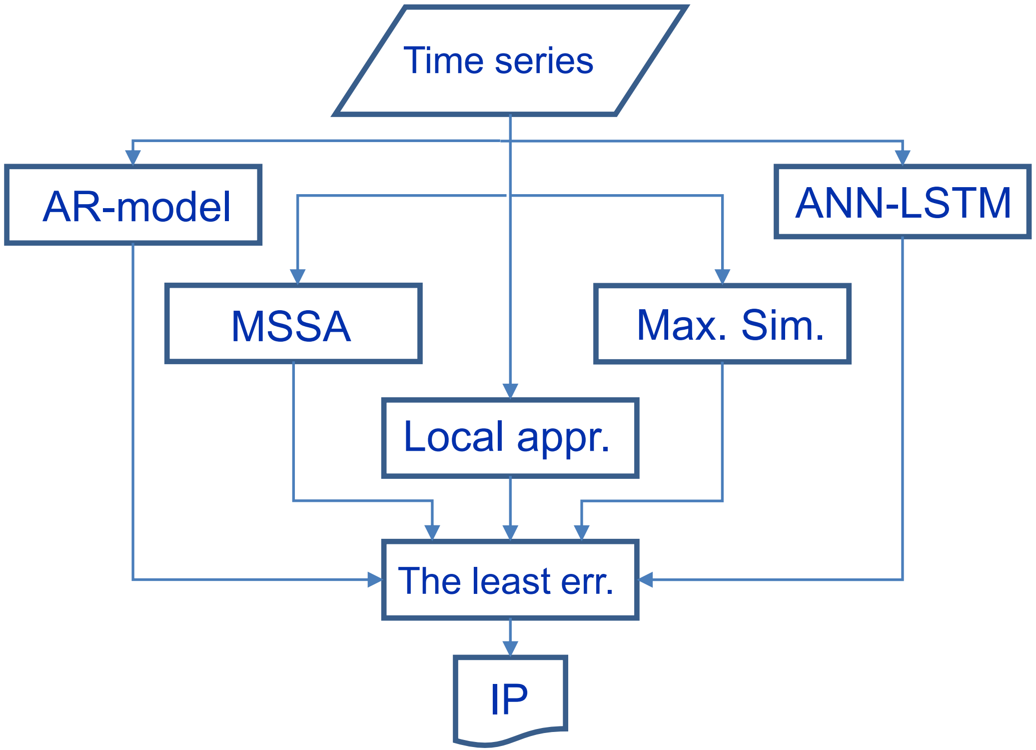
1) To which extent the characteristics of a transition graph are useful for estimating intrinsic predictability (IP)?

2) Is it possible to train an IP forecasting model of a sufficient quality using artificially generated training data?



Intrinsic Predictability

1. Forecasting quality depends on model.
2. Every time we evaluate a prediction, we deal with **realized predictability**.
3. The low quality may be caused by unsuitable model, or by some reason which concerns the time series itself.
4. This reason is so called **intrinsic predictability** (IP), which is model independent property of the series.
5. We can estimate it choosing the best forecasting result of several models.



Unusual models

The Local Approximation (LA)

The Hankel matrix represents the state space. We choose the nearest neighbours in the state space and make a linear regression on them:

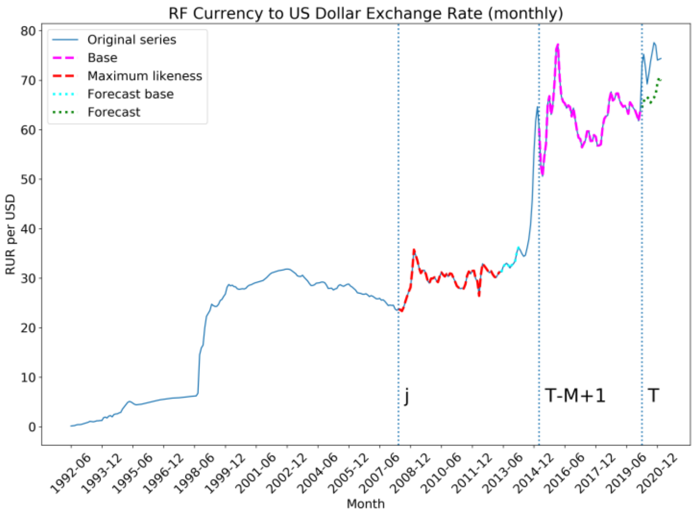
$$\{x_s\}: \sum_{s \in w_m} \left\| x_{N-p+1} - x_s \right\| \to \min_{w_{N-p+1}}$$

$$\{\hat{a}\}: \sum_{w_s} [x_{t+1} - (a_0 + x_t^T \cdot \hat{a})]^2 \to \min_{\hat{a}}$$

A. Y. Loskutov, O. L. Kotlyarov, I. A. Istomin, and D. I. Zhuravlev, “Problems of nonlinear dynamics: Iii. local methods of time series forecasting,” Moscow University Physics Bulletin, vol. 57, 03 2002.

The Maximal Similarity (MaxS)

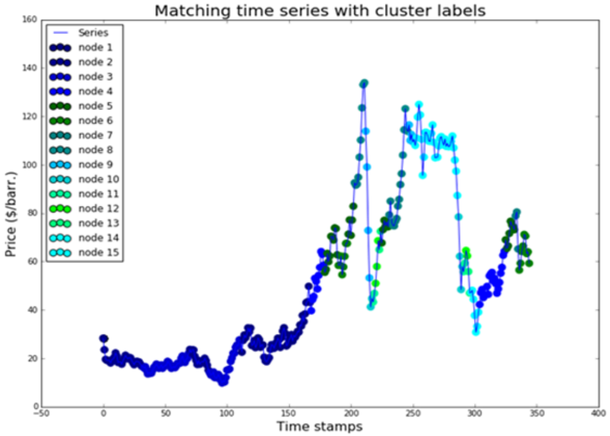
According to the Dirichlet principle, for a pseudo random process with a finite number of internal states the values of the time series will sooner or later be repeated.



I. Chuchueva, “The time series extrapolation model based on maximum likeness set,” Science and Education of the Bauman MSTU, no. 1, 2010.

Time Series and Transition Graph

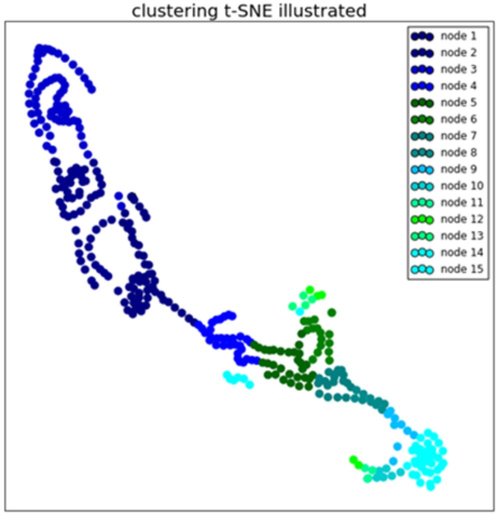
УНИВЕРСИТЕТ ИТМО



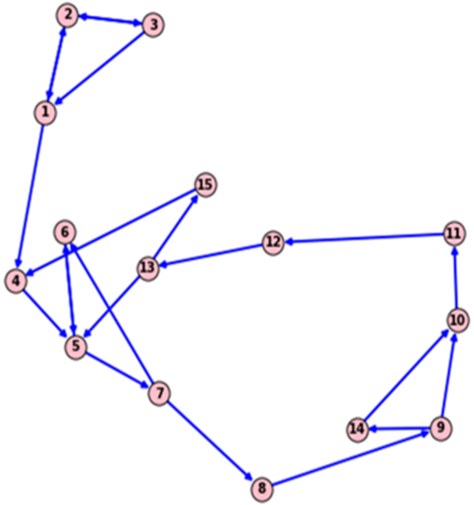
$$\{x_1, x_2, x_3, \cdots, x_N\} \rightarrow$$

$$\boldsymbol{X}_{p\times(N-p+1)} = \begin{pmatrix} x_p & \cdots & x_N \\ \vdots & \ddots & \vdots \\ x_1 & \cdots & x_{N-p+1} \end{pmatrix}$$

Hankel matrix represents the state space of the system *(Takens theorem)* *Whitney embedding theorem* helps to calculate the depth of these delays *(p)*



Points in the state space are clustered by distances. These clusters are the vertices of the graph



Graph edges are the system transitions from one node to another according to the time series

Transition Graph Characteristics

USEFUL FOR CLUSTERING

Graph entropy is a function depending both on the graph itself and on a probability distribution on its vertex set.

$$E(G) = \sum_{i=1}^{n} D_i(G) \log_2 \frac{1}{D_i(G)}$$

D_i is degree centrality for each of the graph vertices.

Connectivity counts the minimum number of elements (vertices or edges) that need to be removed to separate the remaining vertices into two or more isolated sub-graphs.

Normalized number of short cycles can be calculated the ratio of the loops that include less than $\frac{1}{5}$ of all vertices with respect to all cycles in the graph.

USELESS

Density is defined to be the ratio of the number of edges with respect to the maximum possible edges.

$$D(G) = \frac{|E(G)|}{|V(G)| \cdot (|V(G)| - 1)}$$

Modularity is a measure of the structure of networks or graphs which measures the strength of division of a network into modules.

Assortativity is a preference for a graph vertices to attach to others that are similar in some way. The assortativity coefficient is the Pearson correlation coefficient of degree between pairs of linked vertices

Time Series Features

Kolmogorov-Sinai entropy is the supremum of the entropy rate $h_\mu(T, \xi)$ over all finite partitions.

$$h_\mu^{KS}(T) = \sup_{\xi} h_\mu(T, C)$$

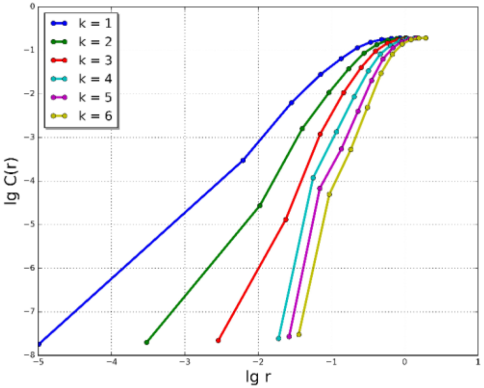
$$h_\mu(T,\xi) = -\lim_{n\to\infty}\frac{1}{n}\sum_{i_1,\ldots,i_n}\mu(T,C)\times ln\,\mu(T,C)$$

Correlation dimension.

$$C(r) = \sum_{i=1}^{m}\sum_{j=i+1}^{m}\frac{\theta\big(r-\rho(i,j)\big)}{m(m-1)}, \quad \theta(x) = \begin{cases} 1 if x > 0 \\ 0 if x \le 0 \end{cases}$$

$$\rho_k(i,j) = \sqrt{\sum_{l=1}^{k}\left(x_{i-k+l} - x_{j-k+l}\right)^2}$$

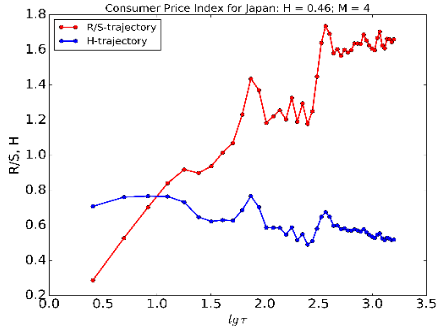
$$d_k = \lim_{r\to 0}\lim_{m\to\infty}\frac{lnC(r)}{lnr}$$



Hurst exponent.

$$H(\tau) = \lim_{\tau\to\infty}\frac{ln\frac{R(\tau)}{S(\tau)}}{ln(\alpha\tau)} \qquad S(\tau) = \sqrt{\frac{1}{\tau}\sum_{t=1}^{\tau}(x_t - \acute{x}_\tau)^2}$$

$$R(\tau) = \max_{1\le t\le\tau}\sum_{j=1}^{\tau}(x_j - \acute{x}_\tau) - \min_{1\le t\le\tau}\sum_{j=1}^{\tau}(x_j - \acute{x}_\tau)$$



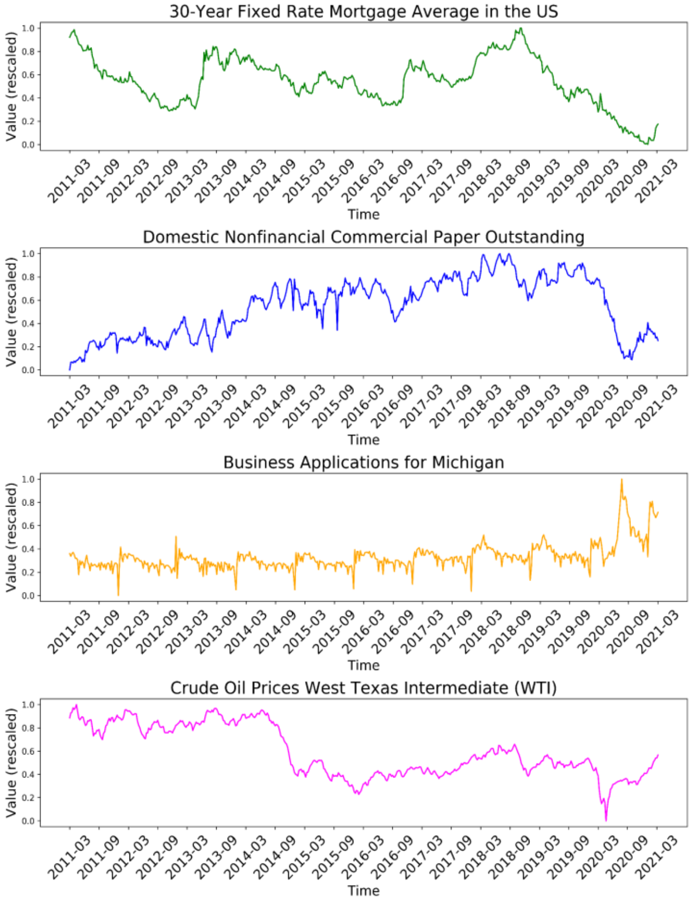
Noise measure and Random walk detection

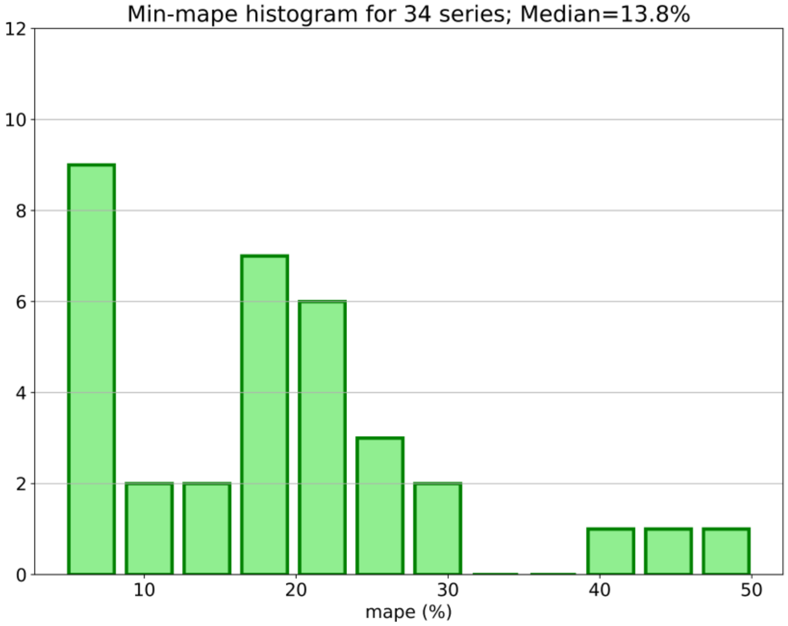
$$F_N = 1 - \sqrt{\frac{N\cdot\sum_{i=1}^{N-1}(x'_i - \acute{x}')^2}{(N-1)\cdot\sum_{i=1}^{N}(x_i - \acute{x})^2}}$$

ITMOre than a UNIVERSITY

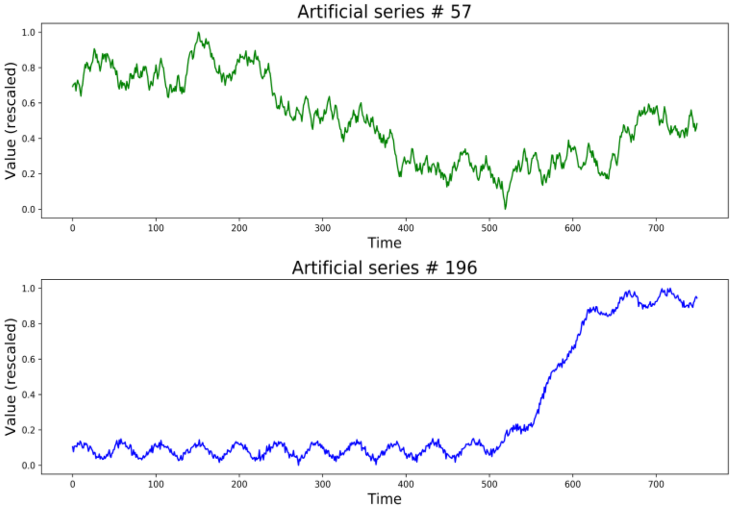
Data Sets

Real world (34 series)

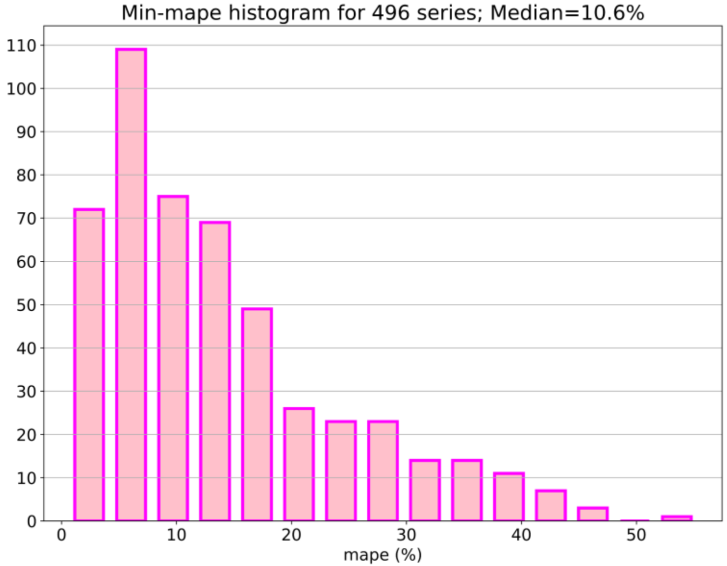




Data Sets



Artificial (496 series)



$$x_t = \mu_1 x_t^{per} + \mu_2 x_t^{tr} + \mu_3 x_t^{noise} + \mu_4 x_t^{brown}$$

$$x_t^{per} = \sin\frac{1}{15}t \qquad x_t^{tr} = 1 - \frac{1}{1+\exp(-75+\frac{t}{25})}$$

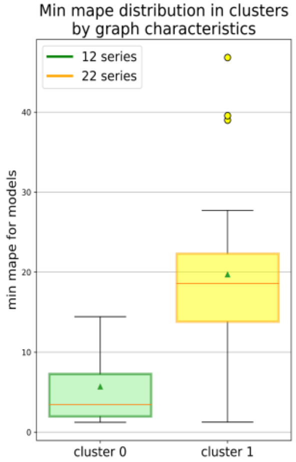
$$x_t^{noise} = \mathcal{N}(1,1) \qquad x_t^{brown} = x_{t-1}^{brown} + \mathcal{N}(1,1); x_0^{brown} = 0$$

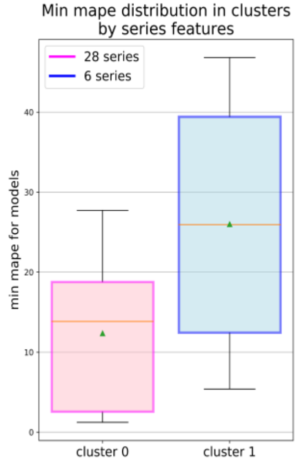
$\mathcal{N}(1,1)$ – normally distributed random values with mean 1 and standard deviation 1

Experiments with Clustering

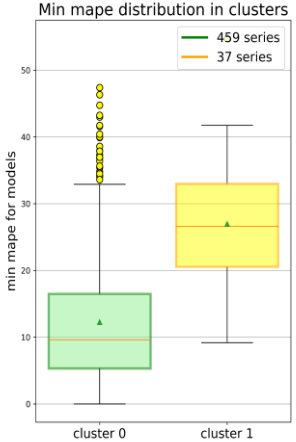
УНИВЕРСИТЕТ ИТМО

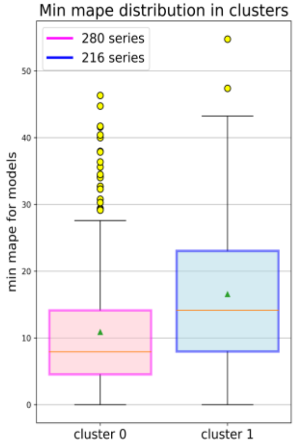
Real-world series clustering



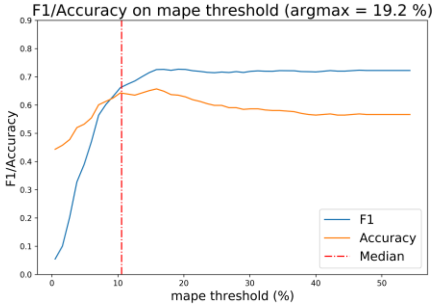


Artificial series clustering

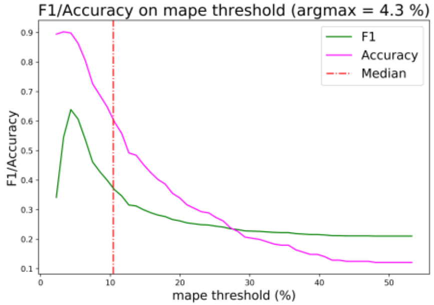




For graph characteristics



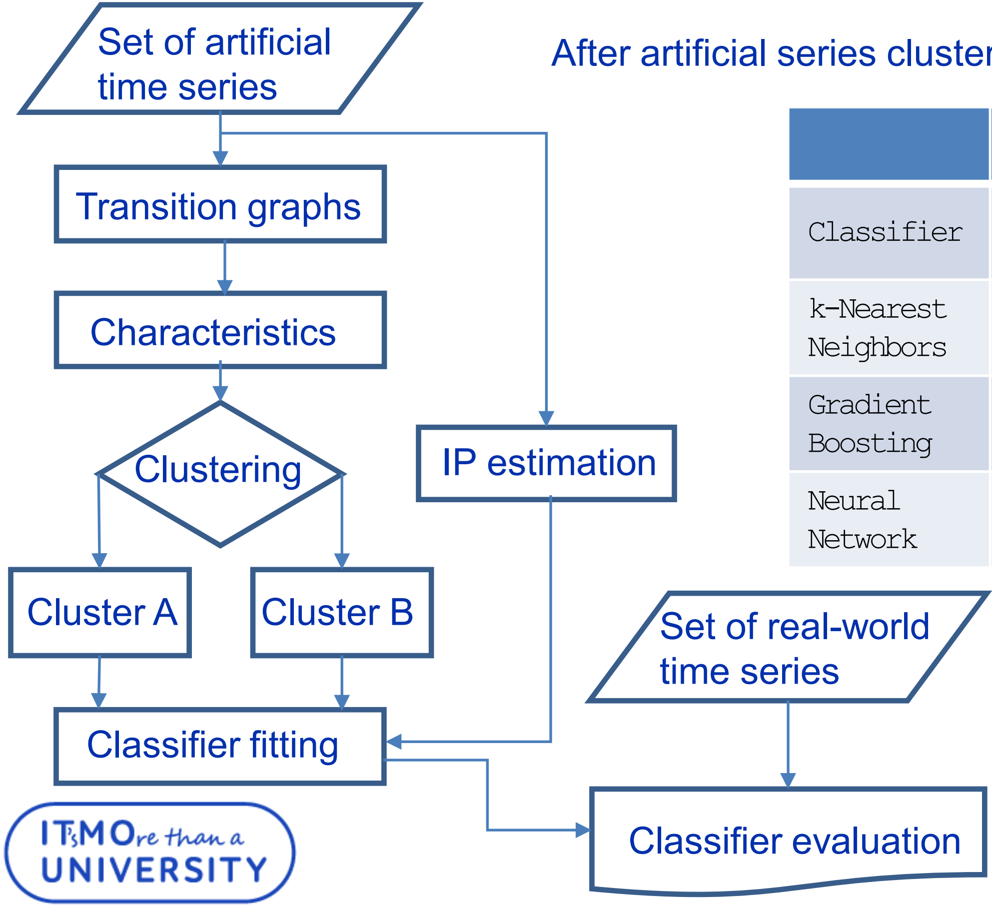
For time-series features



Mann-Whitney U-test

	Transition graph		Time series	
Series	*U-score*	*p-value*	*U-score*	*p-value*
Real	36	0.0004	52	0.0773
Art.	436	$2.8 \cdot 10^{-6}$	2805	$2.3 \cdot 10^{-5}$

Experiments with Classifier Fitting



After artificial series clustering three classifiers were fitted…

	Graph method			Series method		
Classifier	Accur. test	Accur. real	F_1	Accur. test	Accur. real	F_1
k-Nearest Neighbors	78.66%	73.53%	0.74	81.10%	73.53%	0.74
Gradient Boosting	82.93%	76.47%	0.87	80.49%	73.53%	0.85
Neural Network	84.76%	76.47%	0.87	85.98%	76.47%	0.87

… and applied for Real-world series classification

Conclusions and Future Work

1. The presented method of (unsupervised) dividing time series into two types (with “good” and “bad” predictability) showed that the use of transition graph characteristics may significantly improve the procedure of estimating the expected time series forecasting quality.
2. Compared to the procedure using the features of initial time series, the graph method provides more separated clusters. While the quality of graph and series methods is barely differing, graphs can be more useful when the task is to select the series of “good” predictability.
3. We also plan to explore the dynamics of time-dependent process predictability by means of graph representation and incremental machine learning methods. We suppose that the dynamically growing graph can provide some valuable information on critical transitions in dynamic systems.

This research is financially supported by The Russian Science Foundation, Agreement № 17-71-30029 with co-financing of Bank Saint Petersburg.

Best regards!

www.ifmo.ru

ITsMOre than a UNIVERSITY