Abstracts of the Ninth Workshop on Simulation (2018) 1-1

Model-free classification of panel data via the ϵ-complexity theory

Alexandra Piryatinska[1], Boris Darkhovsky [2]

Success in solution of classification problem essentially depends on a right choice of a feature space. In majority of applications features spaces are constructed based on prior information on data generating mechanisms. However, such information is not always available. Typical example here is an electroencephalogram signal. According to most of experts, there are no generally accepted models of such signals and therefore it is difficult to analyze them. The same situation is typical in applications to financial time series, some biological problems, etc., where also there are no standard models. This circumstance significantly complicates creation of feature spaces for classification problems.

In this talk, we consider the problem of classification for so called *panel data*. In statistics and econometrics, panel data are multi-dimensional data involving measurements over time. Typically a problem to identify structural break in panel data is considered in the literature. For example, such phenomena may be produced by governmental policy changes, the introduction of a new technology, etc. We will classify different groups of simulated panel data which are generated by different mechanism. We use multidimensional linear model to simulate panel data. Such models are typical in econometrics. Different classes of the data generated by different matrices of states and noises. Such classification can be used further for detecting the moments of time at which these changes occurrences.

To perform such classification, we employ our theory of the *ϵ-complexity* of continuous map, in particular, continuous vector-functions. In the majority of applications, we deal with vector-functions given by their values at a discrete set of points (i.e., by a finite sample of finite-dimensional vectors). We will assume that this set of values is *the restriction of a continuous vector-functions on some uniform grid*. In our talk we generalized the main result of our theory to the case when a continuous map is given by its values on some uniform grid. We show that in this case the ϵ-complexity can be effectively characterized by two real numbers - *the complexity coefficients* - for "almost all" Hölder maps.

We describe our algorithm for calculation of the complexity coefficients and show that these coefficients can be used as features for classification of multivariate time series data including panel data. Therefore, complexity coefficients can be used also for detecting change-points in samples of panel data.

The results of simulations will be presented.

[1]San Francisco State University, San Francisco, USA E-mail: alpiryat@sfsu.edu

[2]Institute for Systems Analysis FRC CSC RAS, Moscow, Russia E-mail: darbor2004@mail.ru