ADAPTIVE METHODS OF SOLVING INVERSE PROBLEMS IN SPECTROSCOPY OF NATURAL WATERS*

Sergey Dolenko

D.V.Skobeltsyn Institute of Nuclear Physics, M.V.Lomonosov Moscow State University, Moscow, Russia; dolenko(at)srd.sinp.msu.ru

A wide family of machine learning algorithms are often called adaptive or data-driven methods, due to their capability of adaptation to the available set of data, learning by example, requiring no physically grounded analytical or computational model or *a priori* knowledge of the studied object. The scope of problems that are solved by such methods includes those of prediction, evaluation, classification, clusterization, inverse problems, and other data analysis problems. Examples of such methods are artificial neural networks (ANN) and the method of partial least squares, or projection to latent structures (PLS). From mathematical point of view, these methods are sophisticated approximation methods using adaptively tuned combinations of relatively simple functions of most general type.

As other physical methods based on indirect measurements, laser spectroscopy implies solution of inverse problems (IP) – determination of the values of the sought-for parameters by the shape of measured spectra. It should be noted that spectra of aqueous media, including natural waters, are especially difficult to model, due to possible presence of multiple dissolved components, and their interaction among each other and with hydrogen bonds of water molecules. That is why adaptive methods of IP solving based on approximation of the inverse function are demanded and can be efficiently used in laser spectroscopy of natural waters.

Attention is driven to the main differences between ANN and PLS, and the main shortcoming of PLS – that it is a linear method (yet it is the best linear method). Even with an adequate non-linear pre-processing of data, PLS is often unable to build an approximation comparable by its quality with that implemented by an ANN. The main advantage of PLS is its low computational cost.

In the lecture, methodological aspects of using ANN are discussed. From the point of view of data processing methods, any IP can have various formulations: as a regression, classification (for discrete-valued IP) or optimization problem. The key differences of ANN as a method of solving IP from alternative methods are discussed.

When solving IP, ANN can be used within one of several methodological approaches: “model-based”, “experiment-based”, and “quasi-model”. The difference among these approaches, their properties and areas of application are described.

A separate question arises if the IP being solved is a multi-parameter one. The possible approaches to the order of determination of parameters are autonomous determination, simultaneous determination of all parameters, group determination (with joining of parameters into groups with simultaneous determination within each group), and stepwise determination (when some of the parameters already determined are used as additional inputs for determination of other parameters).

The material is illustrated at the example especially interesting for the specialists in ecology and optics of coastal zones. The discussed IP is that of identification and determination of concentrations of inorganic ions in multi-component aqueous solutions by their Raman spectra, using ANN or PLS. The process of IP solution is discussed step by step, from problem statement and analysis of spectra properties through selection of the methodological approach to additional

* This study has been performed at the expense of the grant of Russian Science Foundation (project no.14-11-00579).

methods such as group determination, cluster-based approach, and training with noise. It is stressed that with increasing number of components in the multi-component solution, the complexity of the problem increases, and linear methods begin to fail; ANN turn out to be more resilient to this increasing complexity.

The general purpose of the lecture is to attract attention of a wide audience of spectroscopists working with natural waters and with aqueous media in general to the great opportunities opened by use of adaptive methods and of the latest methodologic achievements in IP solution by ANN. The lecture is to be followed by practical exercise demonstrating solution of classification problems by ANN, the importance of data representativity, and the correct methods of ANN training.