Neural Network Solution of Inverse Problems of Spectroscopy with Integration of Various Types of Spectra: Iterative Feature Selection with Accounting for Redundancy

Nickolay O. Shchurov, 1, 2, Igor V. Isaev, 2, 3, Olga E. Sarmanova, 1, 2, Sergey A. Burikov, 1, 2, Tatiana A. Dolenko, 1,2, Kirill A. Laptinskiy, 2, Sergey A. Dolenko, 2

1 Faculty of Physics, Moscow State University, Moscow, Russia
2 D.V. Skobeltsyn Institute of Nuclear Physics, M.V. Lomonosov Moscow State University, Moscow, Russia
3 Kotelnikov Institute of Radioengineering and Electronics, Russian Academy of Sciences, Moscow, Russia

In the neural network (NN) solution of many physical problems, it becomes necessary to reduce the dimension of the input data in order to achieve a more accurate and stable solution while reducing computational complexity. When solving the inverse problem of spectroscopy, high multicollinearity between input features (IF) is often observed, since spectral lines can be much wider than the width of the spectral channel. This requires use of a feature selection method that takes into account this property.

The method used in this study is based on iterative selection of IF with the highest Pearson correlation with the target variable and exclusion of IF with high cross-correlation. We analyze the quality of the NN solution of the inverse problem of determining the concentrations of heavy metal ions in water by Raman scattering, IR absorption and optical absorption spectra. Integration of various types of spectroscopy may improve the results of IP solution even with no feature selection, as different types of spectra may carry different information about the object.

In this study we demonstrate the efficiency of combined use of the iterative selection of IF and of integration of various types of spectroscopy. The results are shown on the complete set of features and on its subsets obtained using the method under consideration.

The research was carried out at the expense of the grant of the Russian Science Foundation No. 19-11-00333, https://rscf.ru/en/project/19-11-00333/.