UDC 620.193.4, 004.81

DEVELOPMENT OF AN APPROACH FOR GENERATING CYCLIC VOLT-AMPERE CHARACTERISTICS OF CORROSION INHIBITION PROCESSES BASED ON EXPERIMENTAL DATA

Krotkov N.A. (ITMO)
Scientific supervisor – assistant of the Scientific and Educational Centre of Infochemistry, Aliev T.A. (ITMO)

Introduction. Modern electrochemistry requires complex multicomponent systems investigation. However, conventional experimental methods can yield noisy and shifted cyclic voltammetry (CVA) signals. For example, corrosion processes in flow systems during oil transportation [1] and the analysis of antibiotics in milk [2] are complicated by three or more components, which are significantly difficult for signal interpretation. Machine learning, specifically digital twins, is a promising direction for modeling the behavior of an *in-silico* system, which will reduce the need for expensive and time-consuming experiments. However, the limited amount of data due to the complexity of conducting experiments creates difficulties in developing digital doubles. To overcome these problems, a generative approach capable of synthesizing a given number of CVA curves is proposed, provided that the raw data is carefully cleaned, calibrated, and filtered beforehand, minimizing the effects of noise and bias. The additional use of methods that consider the specific characteristics of electrochemical processes – namely, quantum-mechanical descriptors and QSAR descriptors – will enable the generation of data that reflect the true physicochemical patterns, as well as the adaptation of the model for various systems, particularly in the determination of antibiotics in milk and the processes of corrosion inhibition in pipelines.

Main part. A combined architecture of a Variational Autoencoder (VAE) and a Generative Adversarial Network (GAN) was chosen to create the model for generating synthetic CVA characteristics. On the one hand, VAE allows us to compress the feature space (in this case, CVA features, quantum mechanical, and QSAR descriptors) into a more compact latent representation, leaving the most relevant features. On the other hand, VAE can generate new data based on the hidden representation [3]. One of the main advantages of the proposed architecture is that it enables adversarial training: the VAE learns to generate new data that are as similar as possible to real data, while the discriminator learns to distinguish real data from the generated data. A high level of noise and signal shift in CVA greatly limits the use of statistical data processing to build calibration curves. However, the application of machine learning makes it possible to work effectively with such systems, as demonstrated in paper [2], where the authors detected antibiotics in milk. The model was trained on CVA characteristics and achieved high accuracy in classifying antibiotics in milk.

The first step in constructing a generative model is data preparation for training. Characteristics of CVA were obtained at different concentrations of four corrosion inhibitors used as data. Procedures were used to process these characteristics: interpolation to ensure the same length of CVA, elimination of extreme curves that deviate from the normal distribution, and removal of uninformative data points. We used the density functional theory (DFT) to obtain the energy of the HOMO-LUMO transition and other characteristics that describe CVA. In addition, QSAR descriptors were generated using the RDKit library. The processed data was divided into training and validation samples in the 80/20 ratio and fed to the input of a generative model, which was implemented using the Python programming language and the PyTorch framework.

Conclusion. The model VEG-GAN was developed and trained using experimental data (CVA characteristics) and theoretical descriptions (DFT and RDKit). The resulting model generates high-quality synthetic data identical to experimental data. Thus, the obtained results open the way to establishing a digital twin of electrochemical systems.

References:

1. Tyusenkov A. S., Nasibullina O. A. Corrosion of tubing of oil fields // IOP Conference Series: Materials Science and Engineering. – 2019. – Vol. 687, No. 6. – 066016. – doi: 10.1088/1757-899X/687/6/066016.

2. Aliev T. A. et al. Electrochemical Sensor to Detect Antibiotics in Milk Based on Machine Learning Algorithms //ACS Applied Materials & Interfaces. – 2023. – T. 15. – №. 44. – P. 52010-52020

3. Larsen A. B. L. et al. Autoencoding beyond pixels using a learned similarity metric //International conference on machine learning. – PMLR, 2016. – P. 1558-1566.