UDK 54.061, 54.062

DATA PROCESSING AND MACHINE LEARNING FOR SPECIALLY DESIGNED REACTION-DIFFUSION SYSTEM ANALYSIS

Aliev T.A. (ITMO University), **Timralieva A.A.** (ITMO University)
Scientific supervisor – Dr., Prof. Skorb E.V.
(**ITMO University**)

Reaction-diffusion system is an efficient tool to form capsules for DNA storage, but its analysis is difficult due to its special design. We combine machine learning approach, e.g. pretrained convolutional neural networks, and data processing tools to define both conditions of capsules' formation and its quantitative characteristics.

Capsule's formation is an essential stage for active substances storage and transfer. For DNA we suppose supramolecular melamine-cyanurate capsules formation based on reaction-diffusion system. We design agar gel matrixes as triangles consisting of melamine (M), DNA and cyanuric acid (CA) accordingly. Diffusing fronts of M and CA go through DNA containing gel, and the M-CA-DNA capsules form. Due to the space between the diffusing fronts the size of the capsules and its amount is different. The capsule's formation is sensitive to any condition changes, so we vary the concentration of cations to explore the ion regulation of the process and DNA concentration. Not only the size of capsules but also the fluorescence intensity differs.

As all the mentioned parameters of the capsules and its behaviour in the reaction-diffusion system is extremely weakly pronounced and almost similar by sight we use machine learning algorithms such as convolutional neural networks (CNN). The advantage of CNN is an efficient pattern recognition that makes it the best tool for image analysis such as fluorescent microscopy images.

We use pretrained CNN with frozen trained layers and change only the last active layer training it on our dataset. Three different CNNs were used: ResNet50, VGG16, VGG19. The most effective one in our case is VGG16 with ninety-six percent test accuracy. No overfitting is observed, and the confusion matrix shows excellent distribution of right recognition at all in contrast to other CNNs. The graphic interface for recognition of formation conditions is also performed to simplify the analysis.

The dataset enables to get statistical data about capsules' parameters in different sections of reaction-diffusion system under any formation conditions. We use data processing tools to count the particles and get fluorescence histograms automatically to combine it with CNN.

Thus, we get a new analysis system consisting of machine learning and data processing to get not only the formation conditions of the capsules in reaction-diffusion system but also the quantitative characteristics. It is a promising way to get insights in the behaviour of specially designed reaction-diffusion systems in chemistry and biochemistry.

Aliev T.A. (author)

Skorb E.V. (scientific suprevisor)