Deep Learning for Exploration of Chemical Space and Drug Discovery

Project overview:

Machine learning has emerged as a key computing technology in life sciences, and computer-aided drug discovery is one of the fields in the renaissance. Indeed, the number of possible chemical compounds is enormously large (~10^60 compounds), however only minor part (~10^8 compounds), discovered by means of classical combinatorics and manual design, is used in chemistry. Lack of novel chemical scaffolds hinders the drug discovery and drug development processes, while safer and more selective drugs are still in high demand. The breakthrough is expected in applying machine learning methods to accumulated data from experimental biophysical assays in order to derive powerful prediction models for *de novo* molecular design.

PhD objectives:

Recent advances in generative adversarial network (GAN) development allowed to create a chemical generators that produce valid chemical compounds using 1D string representations, as known as SMILE strings. Implementation and optimization of such GAN will be a starting point for the project. Then we will try different chemical representations and various reward functions aiming to generate chemical compounds with relevant pharmacological and physicochemical properties. As the proof-of-concept step we will apply the developed methods along with the virtual ligand screening pipeline in order to generate novel chemicals that binds relevant pharmacological molecules, such as G protein-coupled receptors.

Requirements:

We are looking for creative, passionate and hard-working graduates from all over the world - whether you hold a M.Sc. degree in computer science, computational biology, chemistry, physics, or a related field, as long as you have a passion for working at the interface of data science and biology, this may be the place for you.
The ideal candidate has i) solid background in molecular modelling of protein complexes or computational chemistry, ii) programming experience in C/C++/Python, iii) experience in machine learning and data analysis, and iv) excellent communication skills in English, and should be highly motivated to develop state-of-the-art software and conduct research for open challenges in molecular modelling and computer-aided drug discovery.

Contact:

Please send your application with the keyword ‘PhD application’ as one single file including a CV, a motivation letter, copies of university transcripts (BSc and MSc, in Russian or English), and contact addresses of two academic references by email to p.popov@skoltech.ru.

Research Environment:

PhD project will be supervised by Petr Popov at Skolkovo Institute of Science and Technology in the Center for Computational and Data-Intensive Science and Engineering. PhD fellowship : 75000 rub.

Skolkovo Institute of Science and Technology is a private graduate research institute in Moscow, Russia. Established in 2011 in collaboration with MIT, Skoltech cultivates a new generation of researchers and entrepreneurs, promotes advanced scientific knowledge and fosters innovative technology to address critical issues facing Russia and the world. Skoltech applies the best Russian and international research and educational practices, with particular emphasis on entrepreneurship and innovation.

Skoltech Center for Computational and Data-Intensive Science and Engineering aims at the conduction of cross-cutting interdisciplinary research driven by modern applications in the fields of computational and data sciences. CDISE has accumulated several research groups with the expertise unique in Russia and competitive at the world level in their prospective areas.