Autonomous Decision Making in Automated Cell Profiling for Drug Discovery Applications

María Andreína Francisco Rodríguez (maria.andreina.francisco@it.uu.se), Department of Information Technology, Uppsala University, SE
Ola Spjuth (ola.spjuth@farmbio.uu.se)**,** Department of Pharmaceutical Biosciences and Science for Life Laboratory, Uppsala University, SE
Justin Pearson (justin.pearson@it.uu.se), Department of Information Technology, Uppsala University, SE

Project Aims

The overarching aim of this project is to contribute to the theory and applications of intelligent decision making applied to drug discovery problems. Building on established mathematical theories of confidence predictions, you will develop new artificial intelligence methods that are able to report explanations or valid confidence measures, which you will then combine with various optimisation techniques as a basis for decision making. From an application point of view, the aim is to integrate the developed methods for decision making to control a fully automated biological lab with applications in drug discovery. Given a goal, (for example, identifying the optimal drug combination for a specific cancer tumor), the system should make intelligent decisions for selecting the next set of actions (experiments to carry out), in such a way that it maximizes the utility of available resources (compounds, equipment, etc), while iteratively improving its confidence estimates.

Background

In drug discovery, various types of data are considered in the decision making process; e.g. what drug candidates should be advanced, what experiments should be performed, how should chemical structure be optimized etc. The data is, in many cases, heterogeneous, and it is often fused with data from previous studies, as well as with public and private databases. Contemporary biological experimentation with automated (robotized) laboratories is, in almost all cases, focused on processing a large number of samples in batches, which generates large data resources for retrospective analysis with the hope that it can be used in the future for predictive purposes. For scientists with access to automated labs, the informatics and analytics efforts commonly require substantial manual operations between iterations, making the scientific process inefficient and leading to automated labs being idle, waiting for analyses to complete before proceeding with the next batch of experiments.

Artificial intelligence (AI) techniques are widely used in drug discovery to make predictions on e.g. efficacy and toxicity, but very few examples exist where predictions are used as part of integrated physical systems to guide experimentation. Further, many contemporary AI approaches do not provide explanations, report valid confidence measures, nor class

probabilities for individual predictions, and are often sensitive to the unbalanced datasets that are common in drug discovery.

Project Description, Interdisciplinary Research

We will develop new methods for decision making in iterative drug discovery that, given a specific goal, are able to synthesize the best next action we need to take in order to minimize the cost of reaching the goal. The methods must take into account complicating matters such as e.g. what if the next best experiment requires ordering a new compound? What can be done while waiting for it to arrive? How does experiment batch size affect uncertainty? The limited data sizes available and the relatively high cost and time to generate new data also have implications on the suitability of the methodologies. Here is where we will explore the interactions between mathematical models, optimisation, and various AI techniques to make plans over a long-term horizon.

We will extend the work on confidence-based prediction intervals or probabilities rather than point predictions (conformal prediction) [1], which can cope with very large and unbalanced datasets. Within this learning framework, predictions are hedged and incorporate a valid indication of their own accuracy and reliability. There has been some initial work on active learning with conformal prediction [2], and also conformal prediction for decision making [3], however more theoretical work is needed for combining and expanding these methods and applying them to the problem of decision making in automated robotized biological labs, where we have high degrees of uncertainty about the outcomes of each decision. Finally, we will use the predictions as inputs to a mathematical optimization model for batch selection in active learning. Such models will be designed taking into account particular application characteristics such as long term goals, as well as physical requirements and constraints into the process.

The newly developed methods will have high potential and relevance for many different types of applications, but during this project will be primarily applied to two concrete problems comprising cell-based experiments in an automated lab: 1) Given a chemical compound C and a set of predictive models M_i trained on background data D_i, what set of experiments should be performed next in order to optimize the confidence in a predicted risk assessment of the compound C? 2) Given a library of anticancer drugs and initial data on the response individual drugs have on cells: Select the next experiment consisting of a set of drug combinations that will be executed in the robotized lab so that it maximizes the chances of having a high effect on cancer cells while minimizing the effect (harm) on non-malignant cells. We will also explore and identify combinations of anticancer drugs together with standard drugs (such as painkillers) that show synergistic effects or unexpected toxicities. We will devise and evaluate mathematical optimization models and explore the use of active learning with conformal prediction and optimization using constraint programming (for batch selection) in order to maximize the reward in each round of experiments. For more information on the project application setting, see https://pharmb.io/project/autonomous-phenomics/.

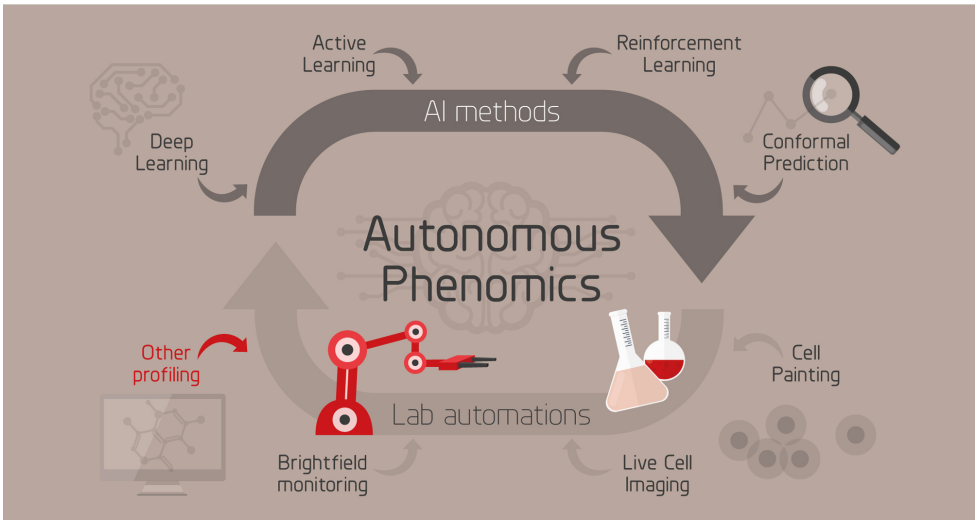

Figure 1: The Autonomous decision making cycle – AI methods can be used to gain a better understanding of experimental results and, in turn, said understanding can then be used to guide future experiments, leading back to more experimental results.

The Research Environment

The main supervisor is Francisco Rodríguez, who is an Assistant Professor in Artificial Intelligence at the Department of Information Technology at Uppsala University. Her current research includes optimization using constraints programming, as well as new computational methods for intelligent decision making with an emphasis on applications in Life Science.

Co-supervisor Spjuth is a Professor at the Department of Pharmaceutical Biosciences and Science for Life Laboratory and head of the Spjuth Lab (https://pharmb.io/) . The lab has established an automated laboratory for cell profiling, and our vision is to develop an autonomous laboratory for cell profiling with applications in pharmacology and toxicology.

Co-supervisor Pearson is a Docent in Computing Science at the Department of Information Technology at Uppsala University. His main research focus is on AI and optimisation.

The applicant will be part of the Optimization Group at the Department of Information Technology and the Pharmaceutical Bioinformatics Research Group (pharmbio) at the Department of Pharmaceutical Biosciences. The position will be co-financed by the Computing Science Division at the Department of Information Technology.

The Successful Candidate

The applicant should have a solid background in applied mathematics, AI, optimization, bioinformatics or a related field. The applicant should have excellent analytical skills, possess curiosity and creativity. Excellent oral and written communication skills in English are required. The applicant should be highly collegial, and be able to work well both in a team as well as independently.

References

1. Vovk V, Gammerman A, Shafer G. Algorithmic learning in a random world. New York: Springer; 2005. p. xvi+324.

2. Matiz S, Barner KE. Conformal Prediction Based Active Learning by Linear Regression Optimization. Neurocomputing. 2020. doi:10.1016/j.neucom.2020.01.018

3. Vovk V, Bendtsen C. Conformal predictive decision making. In: Gammerman A, Vovk V, Luo Z, Smirnov E, Peeters R, editors. Proceedings of the Seventh Workshop on Conformal and Probabilistic Prediction and Applications. PMLR; 2018. pp. 52–62.