1 MACHINE LEARNING APPLICATIONS IN DRUG DISCOVERY

Background

Success in the fields of machine learning and artificial intelligence has led to dramatic improvements in applications towards other areas. One of such is drug discovery, which entails a multi-objective optimization problem spanning several length scales, from solubility (angstroms) to *in vivo* toxicity (meters). A main focus of Acellera is to develop tools that provide understanding and guidance in several structural biology and chemoinformatics tasks.

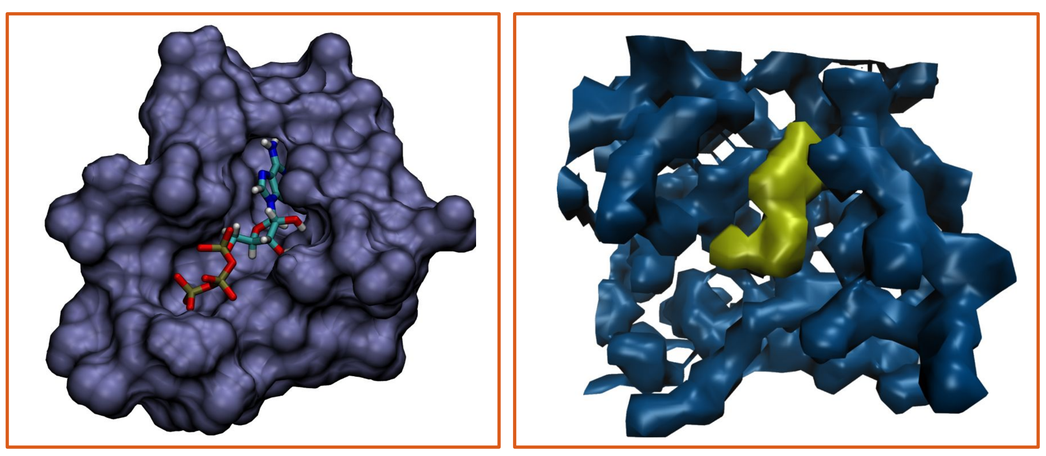

Figure 1. Protein with PDB ID 2HMU bound to ligand ATP (left). Hydrophobic map generated for both protein binding site and ligand, used later in three-dimensional convolutional neural network models (right).

Objectives

The candidate is expected to advance the current state of the art in machine learning applications related to drug discovery and develop those for the free drug discovery platform www.playmolecule.org. Examples include:

- Protein-ligand binding affinity [1]
- *De novo* compound generation [2]
- Protein-protein interface prediction [3]
- Forward and retrosynthesis prediction [4]
- ADMET modeling [5]

Methods

Modern deep learning approaches such as convolutional, recurrent neural networks, as well as generative models such as variational autoencoders and generative adversarial networks are some examples of the models we currently use.

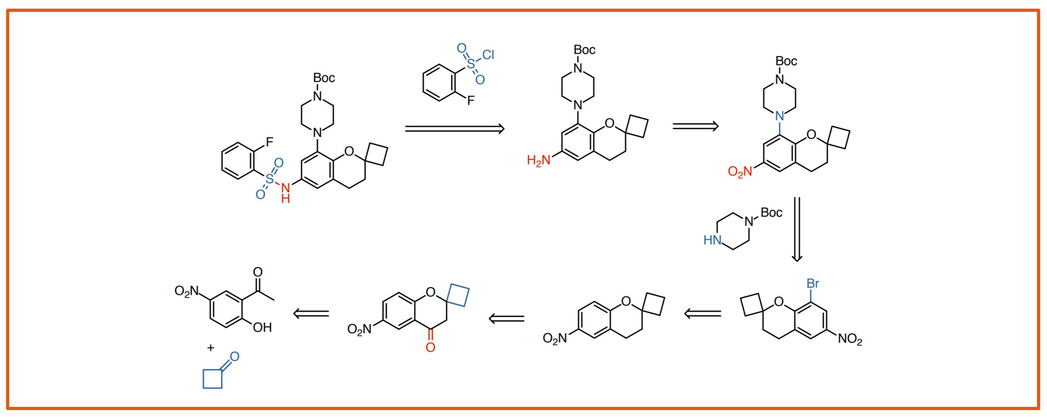

Figure 2. Example of the deconstruction of a ligand's synthesis route via Monte Carlo Tree Search.

Job offer 1

Candidate Profile: The candidate will preferably have a profile in statistics, chemistry, medicinal chemistry or mathematics.

Requirements. Very good communication skills in English. Previous exposure to machine-learning pipelines, Python proficiency, familiarity with Linux and the ability to work with version control systems (e.g. git).

Facilities. Acellera has access to GPU workstations and non-public datasets. Furthermore, it collaborates with several pharmaceutical companies worldwide.

Salary: 1200 euro/month gross.

Future: Possibility of continuing with an industrial PhD.

Location: Acellera labs, C/ Dr Trueta 183, Barcelona

Website: http://www.acellera.com; http://www.playmolecule.org

acellera®

2 FROM MACHINE LEARNING TO STRONG ARTIFICIAL INTELLIGENCE

Background

AI has made significant progress in recent years, reaching superhuman performance on a wide range of tasks. Humans are no longer the best Go players, quiz-show contestants, or even, in some respects, the best doctors. However, state-of-the art AIs cannot compete with simple animals at adapting to unexpected changes in their environment.

AI challenges

Competitions pose several intelligence challenges to assess our best developed AI approaches, showing how far we are from human level intelligence.

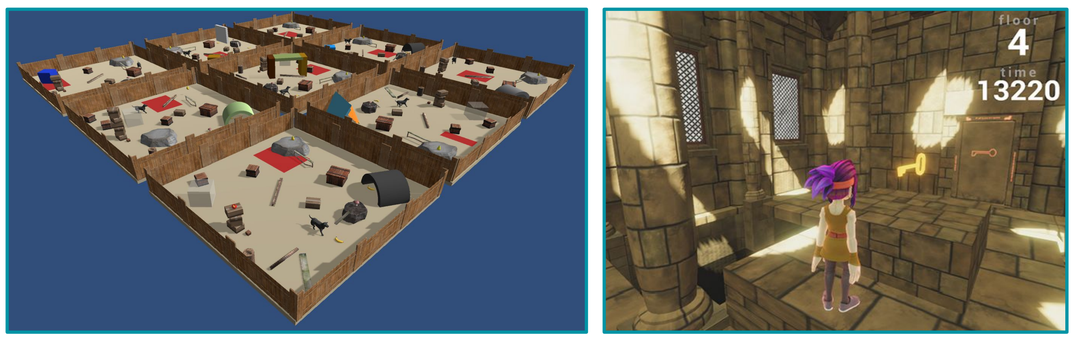

Figure 3. Two examples of challenges. On the left the animal olympics one, where the agent is tasked with tackling problems that some animals can solve. On the right the Google obstacle tower challenge, where an agent needs to navigate through a set of doors, grab keys and solve puzzles.

Methods

Deep reinforcement learning is a popular methodology to solve these tasks, but evolutionary strategies show promise in simpler problems. In the example below:

- the environment is a game simulator, which receives actions and outputs pixel images;
- a model of the world is learned using deep learning methods (a variational autoencoder and a recurrent neural network);
- a controller takes the model representation of the world and predicts which actions produce the highest long term score;
- models are scored on a public leaderboard.

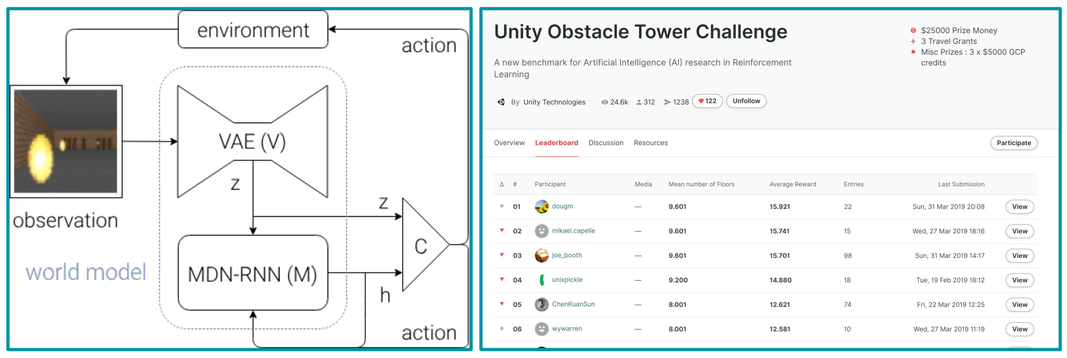

Figure 4. Example framework for solving a 3-dim task (left). The leaderboard of the Google obstacle tower challenge (right).

Job offer 2

Candidate Profile: The candidate will preferably have a profile in computer science, statistics, physics or mathematics.

Requirements. Very good communication skills in English. Previous exposure to machine-learning pipelines, Python proficiency, familiarity with Linux and the ability to work with version control systems (e.g. git).

Facilities. By working in this project, the researcher will have access to state of the art computational resources and large amount of simulation data, which will be crucial for the development and validation of novel computational protocols. The laboratory is very well equipped as well, with access to a local CPU and GPU cluster.

Salary: 1200 euro/month gross.

Future: Possibility of continuing with a PhD.

Location: Computational science laboratory, led by Icrea Prof. Gianni De Fabritiis at PRBB, Barcelona.

Website: http://www.compscience.org

ICREA

upf. Universitat Pompeu Fabra Barcelona

References

[1] Jiménez, J., Skalic, M., Martínez-Rosell, G., & De Fabritiis, G. (2018). K DEEP: Protein–Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. Journal of chemical information and modeling, 58(2), 287-296.

[2] Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., ... & Aspuru-Guzik, A. (2018). Automatic chemical design using a data-driven continuous representation of molecules. ACS central science, 4(2), 268-276.

[3] Fout, A., Byrd, J., Shariat, B., & Ben-Hur, A. (2017). Protein interface prediction using graph convolutional networks. In Advances in Neural Information Processing Systems (pp. 6530-6539).

[4] Segler, M. H., Preuss, M., & Waller, M. P. (2017). Learning to plan chemical syntheses. arXiv preprint arXiv:1708.04202.

[5] Mayr, A., Klambauer, G., Unterthiner, T., & Hochreiter, S. (2016). DeepTox: toxicity prediction using deep learning. Frontiers in Environmental Science, 3, 80.

[6] http://animalaiolympics.com/

[7] https://blogs.unity3d.com/2019/02/18/the-obstacle-tower-challenge-is-live/

[8] David Ha, Jürgen Schmidhuber, World Models, https://arxiv.org/abs/1803.10122; https://worldmodels.github.io/