
Protein Design and Game Theory

PROJECT PROPOSAL FOR MASTER THESIS

Motivation

Protein design is an emerging ambitious goal within the bioengineering industry and academia. In this task, we are interested in discovering an enzyme with desirable properties as quickly as possible. The publicly available datasets are usually not sufficient to train a ML model that is able to predict a desirable protein immediately (one-shot prediction). Usually, one requires a couple of interaction with the nature and synthesize proteins in order to understand what aspects are important which protein variants are viable and which are not [Romero et al., 2013]. The proteins are composed of 20 basic building blocks: amino-acids. A length of a typical protein is around 200 amino-acids, meaning there are 20^{200} conceivable combinations. Of this space very few are active and even less useful, however searching in such big search spaces is challenge for any sequential-decision problems. Usually these problems are specified by defining a utility function that represent the “*information*” a given protein candidate reveals to us one synthesized in laboratory. However, evaluating this utility is infeasible in such large search spaces. Hence, one needs an alternative approach for the utility optimization.

Problem

In this project, we would like to use elements from game theory such as equilibrium learning to address the challenging problem of optimal protein design. Namely, we reduce optimal protein design to the problem of finding an equilibrium of a cooperative game between different amino-acid candidates in different parts of the protein. A goal here would be to formalize this framework, test in simulation and practice and devise solutions to find the appropriate players - parts of the protein that are playing the game.

In particular we are interested in incorporating the novel approaches to optimization and game theory as in [Sessa et al., 2021] and [Sessa et al., 2022]

Contact

This project is best suited for an exceptional student, with prior exposure to reinforcement learning and/or sequential decision-making and game theory. Prior attendance of the class *Probabilistic Artificial Intelligence* is of advantage. If you are interested, please contact Mojmír Mutný (mojmir.mutny@inf.ethz.ch) and Pier Giuseppe Sessa (piergiuseppe.sessa@inf.ethz.ch).

References

[Romero et al., 2013] Romero, P. A., Krause, A., and Arnold, F. H. (2013). Navigating the protein fitness landscape with gaussian processes. *Proceedings of the National Academy of*

Sciences, 110(3):E193–E201.

- [Sessa et al., 2021] Sessa, P. G., Bogunovic, I., Krause, A., and Kamgarpour, M. (2021). Online submodular resource allocation with applications to rebalancing shared mobility systems. In Meila, M. and Zhang, T., editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 9455–9464. PMLR.
- [Sessa et al., 2022] Sessa, P. G., Kamgarpour, M., and Krause, A. (2022). Efficient model-based multi-agent reinforcement learning via optimistic equilibrium computation. In Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., and Sabato, S., editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 19580–19597. PMLR.