IDP:
Stochastic optimization of Biopharmaceutical operations using Reinforcement Learning

Contact: Mirko Schömig (mirko.schoemig@tum.de)

Topic:

Biopharmaceuticals are pharmaceutical drugs derived from biological sources used for therapeutic or diagnostic purposes. They possess a particularly high efficacy and efficiency in treating complex health conditions, like cancer, inflammatory diseases or metabolic disorders.

The primary biological production process is structured into two main stages: The upstream process (USP) contains all production steps related to the cultivation of the living organisms, while the downstream process (DSP) comprises the separation and purification of the active pharmaceutical ingredient (API).

Due to its dependence on biological systems, one has to deal with various operational uncertainties in biopharmaceutical production. In the USP the operations manager needs to decide when to harvest the production batch from the bioreactor, considering the API concentration accumulated up to this point. The formation rate though, is uncertain. In the DSP, chromatography is the most important purification technology. It relies on expensive consumables, which deteriorate with each purification cycle, resulting in a lower purification capacity per cycle. The rate of decay depends on the protein load coming from USP, but due to poor understanding of this impact, we consider the rate of decay to be uncertain. Considering the trade-off of high cost of replacement compared to lower purification capacities per cycle, the operations manager needs to decide when to replace the consumables.
This IDP deals with the integrated optimization of the up- and downstream operations of the primary biopharmaceutical production. The goal is to determine the optimal bioreactor harvesting and chromatography consumable exchange policy, considering the current USP product quantity and DSP purification capacity under uncertain product formation and uncertain chromatography consumables capacity decay.

To deal with this problem we have formulated the problem as a Markov Decision Process (MDP) model. With this, we can incorporate the problem characteristics of discrete decision time points, multi-periodicity and uncertainties. The MDP can either be solved optimally using dynamic programming, or approximately using reinforcement learning. The problem with dynamic programming is the need for discretization of the state variables and the limited problem size that can be solved. Using reinforcement learning one can solve much larger problem instances, closer to actual industry cases.

The objective of this IDP is to solve the already formulated MDP model using reinforcement learning. For this, an appropriate reinforcement learning algorithm needs to be selected and implemented in Python. Subsequently, a numerical case study should be performed to derive general insights on the optimal operations policy.

**Tasks:**

Your task will be to:

1. Review different reinforcement learning approaches and select the most suitable for this MDP model.
2. Implement the MDP model and reinforcement learning algorithm in Python.
3. Perform a numerical study and derive general insights on the optimal operations policy.
Prerequisites:

You should be familiar with Markov decision processes (MDP), reinforcement learning algorithms and their implementation in Python. Insights to the biopharmaceutical industry, specifically the production process, are a plus but not required.