Explore the Protein Folding Problem Using Reinforcement Learning

Lu, Jiachen (School: Culver Academies)

Proteins are critical for lives. They not only build 10%-35% of our body tissues, but also can be used to understand the structures of different viruses, and then help us to explore effective vaccines. Hence, predicting new protein structures is very important for human health. However, the structure of protein is complicated. Exploration using human experiments is cost-consuming. Recently, artificial intelligence (AI) technology, such as imitation learning and reinforcement learning (RL), has been rapidly developed and significantly improved the efficiency in many different domains. In this project, we will try to use RL to solve the protein folding structure prediction problem. First, we adopted the PH structure as a relatively simple representation of the protein structure, where different peptides can be categorized into two types: P(hydrophilic) and H(hydrophobic). The goal of the protein folding is to try to make more H pairs during the folding process. We then formulated the protein folding problem as a reinforcement learning process. If a new H pair is generated during folding, we collect -1 reward. Such RL reward is designed based on the protein dataset (Protein Data Bank). Finally, we implemented three RL algorithms: 1) Q-learning, 2) Deep Q-learning, and 3) Double Deep Q-learning (DDQN). We implemented and compared the three algorithms in terms of their accuracy and efficiency. We found that all three algorithms can accurately predict the structures of simple proteins. As protein structures become more complicated, the DDQN is performing better.

Awards Won: