625FinalReport

hrtang

2023-12-14

Proteins, fundamental to biological systems, perform a diverse range of functions vital for cellular and organism life, such as catalyzing reactions, transferring signals, and immune system (Alberts et al. 2022). Accurately predicting these functions is a significant challenge in biology, bridging the gap between an ever-expanding database of protein sequences and their known functions. Traditional experimental approaches, while effective, are often slow and expensive. In contrast, computational methods, particularly deep learning, offer a promising, cost-effective alternative due to their ability to handle complex datasets and the availability of extensive labeled data.

Deep Neural Networks (DNNs), a subset of Artificial Neural Networks (ANNs), are particularly suited for protein function prediction. They excel in extracting intricate features from basic input data, a crucial aspect when dealing with the vast and complex data of protein sequences. Our study focuses on leveraging these capabilities, using amino acid sequences, without the 3-D structure, as inputs for DNN models. We chose Gene Ontology (GO) terms as labels for protein functions, encompassing three primary domains: molecular function, biological process, and cellular component.

This research is part of a broader effort initiated by the Critical Assessment of Functional Annotation (CAFA) competition, aimed at advancing our understanding of protein functions. By developing models trained on amino acid sequences, we aim to contribute to a deeper understanding of how proteins determine cellular and organ function. This insight is not only fundamental to biology but also has potential applications in drug development and disease therapy.

References Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. 2022. *Molecular Biology of the Cell*. 7th ed. W. W. Norton & Company.