Anti-Virus Autobots: Predicting More Infectious Virus Variants for Pandemic Prevention Through Deep Learning

Tan, Glenda (School: Raffles Institution)
Koay, Tze Erhn (School: Raffles Institution)

More infectious variants of viruses can arise from rapid mutations in their protein sequences, creating new infection waves. These variants can evade one's immune system and even infect vaccinated individuals, lowering vaccine efficacy. Hence, to improve vaccine design, this project proposes a deep learning approach to predict future, more infectious variants from an existing virus (exemplified by SARS-CoV-2). The approach consists of (i) an algorithm which acts as a "virus" attacking a host cell. To increase infectivity, the "virus" mutates itself to bind better to the host cell's receptor. 2 algorithms were attempted to generate the mutations – greedy search and beam search. At each time-step, greedy search selects the best variant with the strongest binding to the host receptor while beam search selects the 10 best variants. The strength of this variant-host binding was then assessed by (ii) Optimus PPIme, a transformer neural network we developed, with a high accuracy of 90%. It was developed with masked language modeling pre-training, sharpness-aware minimization and data augmentation. These enabled Optimus PPIme to generalize well to new, unseen variants. With components (i) and (ii), beam search eventually proposed more infectious variants of stronger binding to the host receptor than greedy search, which was corroborated by 3D-docking. Therefore, this approach can potentially enable researchers to develop vaccines that provide protection against future infectious variants before the variants appear in reality, preventing outbreaks and saving lives.