COMPUTER SCIENCE
SEMINAR

COVID-19 Vaccine Design using Mathematical Linguistics

Dr. Liang Huang
Oregon State University

Abstract: Abstract: To defeat the current COVID-19 pandemic, a messenger RNA (mRNA) vaccine has emerged as a promising approach thanks to its rapid and scalable production and non-infectious and non-integrating properties. However, designing an mRNA sequence to achieve high stability and protein yield remains a challenging problem due to the exponentially large search space (e.g., there are 2.4×10^{632} possible mRNA sequence candidates for the spike protein of SARS-CoV-2). We describe two on-going efforts for this problem, both using linear-time algorithms inspired by my earlier work in natural language parsing. On one hand, the Eterna OpenVaccine project from Stanford Medical School takes a crowd-sourcing approach to let game players all over the world design stable sequences. To evaluate sequence stability (in terms of free energy), they use LinearFold from my group (2019) since its the only linear-time RNA folding algorithm available (which makes it the only one fast enough for COVID-scale genomes). On the other hand, we take a computational approach to directly search for the optimal sequence in this exponentially large space via dynamic programming. It turns out this problem can be reduced to a classical problem in formal language theory and computational linguistics (intersection between CFG and DFA), which can be solved in $O(n^3)$ time, just like lattice parsing for speech. In the end, we can design the optimal mRNA vaccine candidate for SARS-CoV-2 spike protein in just about 10 minutes. To conclude, classical results (dating back to 1960s) from theoretical computer science and mathematical linguistics helped us solve the very challenging and extremely important problem in fighting the COVID-19 pandemic.

Bio: Liang Huang (PhD, Penn, 2008) is an Associate Professor of Computer Science at Oregon State University and Distinguished Scientist at Baidu Research USA. He is a leading theoretical computational linguist, and was recognized at ACL 2008 (Best Paper Award) and ACL 2019 (Keynote Speech), but in recent years he has been more interested in applying his expertise in parsing, translation, and grammar formalisms to biology problems such as RNA folding and RNA design. Since the outbreak of COVID-19, he has shifted his attention to the fight against the virus, which resulted in efficient algorithms for stable mRNA vaccine design, adapted from classical theory and algorithms from mathematical linguistics dating back to the 1960s.

Friday, March 19, 2021, 1:00 pm
https://emory.zoom.us/j/92103915275

COMPUTER SCIENCE
EMORY UNIVERSITY