On Text, DNA and Proteins: Data-Intensive Applications in Computational Biology, and Associated Optimization Problems/Solutions

Hagit Shatkay
Computational Biology and Machine Learning Lab
School of Computing
Queen’s University
Kingston, ON

Abstract:

The genomic era, in which we live since the sequencing of the Human Genome, is characterized by tremendous amounts of biomedical data, including long genomic sequences, proteins, gene expression information, and much text in the form of scientific publications. The abundant (and difficult) biomedical challenges require various ways of utilizing this data, and, in turn, can be formulated as different optimization problems. The talk will present several biomedical applications involving sequence and text data, explore ways of viewing them as optimization problems, and discuss solutions for some of them.