INFERENCE CRITERIA FOR HIDDEN MARKOV MODELS IN BIOLOGICAL SEQUENCE ANALYSIS

Broňa Brejová

Faculty of Mathematics, Physics and Informatics,
Comenius University in Bratislava, Slovakia

Hidden Markov models (HMMs) are used for many tasks in DNA and protein analysis, including gene finding, sequence family modeling, and sequence alignment. In these application areas, we typically assume that the input DNA or protein sequence X was emitted by the model and seek the sequence of hidden states S which might have emitted X. Traditionally, such inference is done by the Viterbi algorithm, which finds sequence S maximizing the joint probability of X and S in the model. Alternatively, posterior decoding chooses the most likely state at every position of X separately, However, many other inference criteria can be formulated, often motivated by a particular application domain. I will discuss several examples of HMM inference criteria, with focus on modeling aspects and computational complexity. We have proved that some criteria lead to NP hard problems, while others can be computed by efficient algorithms.

HMMs were also recently introduced for modeling DNA sequencing data produced by Oxford Nanopore devices. This technology can produce very long sequencing reads but also suffers from high error rate. We propose to sample alternative sequences from the HMM to improve alignment of reads to a reference genome or to each other.

Keywords: Hidden Markov models, inference, computational complexity, sequence alignment.

References:

Brejová, B., Brown, D. G., and Vinař, T. (2007). The most probable annotation problem in HMMs and its application to bioinformatics. *Journal of Computer and System Sciences*, 73(7), 1060–1077.

Nánási, M., Vinař, T., and Brejová, B. (2010). The highest expected reward decoding for HMMs with application to recombination detection. In *Combinatorial Pattern Matching (CPM 2010)*, volume 6129 of *Lecture Notes in Computer Science*, pages 164–176. Springer.

Nánási, M., Vinař, T., and Brejová, B. (2015). Sequence annotation with HMMs: New problems and their complexity. *Information Processing Letters*, 115(6), 635–639.

Rabatin R, Brejová, B., and Vinař, T. (2016). Using Sequence Ensembles for Seeding Alignments of MinION Sequencing Data. arXiv preprint *arXiv:1606.08719*.