Course Summary

Broňa Brejová

December 16, 2021

Probabilistic models

- Hidden Markov models (gene finding, phylogenetic HMMs for conserved elements, profile HMMs for protein families)
- Phylogenetic trees and substitution models
- Stochastic context-free grammars
- Gibbs sampling
- Maximum likelihood method
- Expectation maximization (EM)

Statistical methods

- Statistical significance, E-value, P-value
- Positive selection test
- Linkage disequilibrium, association mapping

Practice in dynamic programming

- Sequence alignment
  (global, local, affine gaps, saving memory)
- Hidden Markov models (Viterbi and forward algorithms)
- Computation on trees
  (parsimony, Felsenstein algorithm for likelihood)
- Mass spectrometry (MS/MS)
- Secondary RNA structure

Other

- Integer linear programming
- deBruijn graphs
- Clustering and classification

How to model real-life problems

- Consider what data are avilable, what are relevant questions
- Formulate as a computer-science problem (e.g. score optimization)
- Probabilistic models often lead to a systematic choice of a scoring scheme
- The resulting problem often NP hard
  - Heuristics, approximation algorithms
  - ILP and other techniques for exact solutions
  - Can we change problem formulation?
- Testing: are computation results relevant in a given domain? (is our formulation sufficiently realistic?)

Ďalšie predmety

- **Strojové učenie** 2-INF-150, Vinař/Boža (ZS, 4P, 6kr)
- **Vybrané partie z dátových štruktúr** 2-INF-237, Kováč (ZS, 4P, 6kr)
- **Seminár z bioinformatiky (1)-(4)** 2-AIN-50[56],25[12] (oba semestre, 2S, 2kr)
- **Manažment dát** 1-DAV-202, Brejová, Vinař, Boža (LS, 1P/2C, 4kr)
- **Genomika** 2-INF-269, Nosek a kol. (LS, 2P/1C, 4kr)
- **Výzvy súčasnej bioinformatiky** 1-BIN-105, Brejová, Vinař (LS, 2S, 2kr)
- http://compbio.fmph.uniba.sk/vyuka/