

Sztochasztikus modellek a bioinformatikában

2/0/0/v/3

Tárgyfelelős: Miklós István

További oktatók:

Statisztikai bevezető: A likelihood függvény, ML becslés, Bayes statisztika, az EM algoritmus. Sztochasztikus generatív nyelvtanok: Rejtett Markov-modellek, sztochasztikus reguláris és környezetfüggetlen nyelvtanok. Algoritmusok nyelvtanokon: Forward-backward, Viterbi, Inside-outside, CYK, Baum-Welch tréning, poszterior valószínűségek számolása. Biológiai alkalmazások: mintázatfelismerés biológiai szekvenciákban, protein másodlagos térszerkezet-predikció, RNS térszerkezet-predikció.

Szubsztitúciók időfolytonos Markov-modellekkel történő leírása. Klasszikus nukleinsav és aminosav szubsztitúciós modellek. Statisztikus szekvenciaillesztés: Beszúrás-törlés (indel) modellek. Indel modellek, mint többszörös rejtett Markov-modellek.

Evolúciós fák. A Kingman koaleszcens. A Markov-lánc Monte-Carlo (MCMC) módszer alapjai. Evolúciós fák vizsgálata Bayesian MCMC-vel. Genomátrendeződések vizsgálata.

Irodalom:

Durbin-Eddy-Krogh-Mitchison: Biological sequence analysis. Cambridge University Press. 1998

Lunter, G.A., Drummond, A., Miklós, I., & Hein, J.: Statistical alignment: recent progress, new applications and challenges in: Nielsen (editor): Statistical Methods in Molecular Evolution. Springer series in Statistics for Biology and Health. Springer Verlag, 2005

Miklós István: Bioinformatikai algoritmusok. In: Iványi Antal (szerkesztő): Informatikai algoritmusok. Eötvös kiadó, 2004

Stochastic models in bioinformatics

2/0/0/v/3

Course coordinator: István Miklós

Other instructors:

Introduction to the statistical methods: The likelihood function, ML estimation, Bayesian statistics, EM algorithm.

Stochastic transformational grammars: Hidden Markov Models, stochastic regular grammars, SCFGs. Algorithms on grammars: Forward, Backward, Viterbi; Inside, Outside, CYK. Baum-Welch training, posterior probabilities, posterior decoding. Biological applications: pattern recognition in sequences, protein structure prediction, RNA structure prediction.

Modelling substitutions with time-continuous Markov models. Classical models for substitutions in nucleic acid and protein sequences. Statistical alignment: modelling insertions and deletions as birth-death processes. Transforming statistical alignment models into pair and multiple HMMs.

Evolutionary trees. The Kingman coalescent. Markov chain Monte Carlo methods for inferring the Bayesian distribution of evolutionary trees.

References:

Durbin-Eddy-Krogh-Mitchison: Biological sequence analysis. Cambridge University Press. 1998

Lunter, G.A., Drummond, A., Miklós, I., & Hein, J.: Statistical alignment: recent progress, new applications and challenges in: Nielsen (editor): Statistical Methods in Molecular Evolution. Springer series in Statistics for Biology and Health. Springer Verlag, 2005

István Miklós: Algorithms in bioinformatics, In: Antal Iványi (editor): Algorithms of Informatics. Monda kiadó, 2007