



第7回 2013年11月22日(金) 15:40~16:30

Model selection in sequence analyses

配列解析におけるモデル選択

Michiaki Hamada / 浜田 道昭

Graduate School of Frontier Sciences, University of Tokyo

東京大学大学院新領域創成科学研究科

Due to the recent advent of next generation sequencing (NGS) technology, a huge amount of `omix' data (e.g., genome, methyrome, transcriptome and metabolome) has been produced, and finding important knowledge from them is an important task. Hidden Markov Models (HMMs) are widely utilized for modeling biological data; however conventional HMMs are not able to learn model structures behind the data. In this study, I introduce a mathematically-sound Hidden Markov Model that can automatically learn model structures, and apply it to model epigenetics data, which enables to annotate chromatin states. Computational experiments with simulated datasets indicates that our method automatically learns appropriate model. In addition, we systematically compare our method with ChromHMM with using three types of real dataset, indicating that our method, in general, estimates more states than ChromHMM. Finally, I will present several variants of the model, with applications in sequence analyses.

Keywords : sequence alignment, chromatin states, probabilistic model, Hidden Markov Model, Model selection, learning algorithm