



Profile HMMs for DNA Sequence Families: the Conditional Baum-Welch and Dynamic Model-Surgery Algorithms

Paul Edlefsen

Dissertation Advisors: Professors Arthur P. Dempster and Jun S. Liu

Profile hidden Markov Models (Profile HMMs) are widely used for protein sequence family modeling, but are rarely used for modeling DNA sequence families because the Baum-Welch EM algorithm used to parameterize Profile HMMs performs particularly poorly in the DNA context. I will report the results of a simulation study comparing the Baum-Welch algorithm to two new approaches, Conditional Baum-Welch and Dynamic Model Surgery, showing that these provide a great improvement over Baum-Welch in both the protein and DNA contexts. I will also compare these methods in the context of the transposon (interspersed repeat) modeling problem that originally inspired the research.