Search or Article-id

All papers

Mathematics > Statistics Theory

A context dependent pair hidden Markov model for statistical alignment

Ana Arribas-Gil, Catherine Matias (SG)

(Submitted on 14 Jul 2011)

This article proposes a novel approach to statistical alignment of nucleotide sequences by introducing a context dependent structure on the substitution process in the underlying evolutionary model. We propose to estimate alignments and context dependent mutation rates relying on the observation of two homologous sequences. The procedure is based on a generalized pair-hidden Markov structure, where conditional on the alignment path, the nucleotide sequences follow a Markov distribution. We use a stochastic approximation expectation maximization (saem) algorithm to give accurate estimators of parameters and alignments. We provide results both on simulated data and vertebrate genomes, which are known to have a high mutation rate from CG dinucleotide. In particular, we establish that the method improves the accuracy of the alignment of a human pseudogene and its functional gene.

Subjects: **Statistics Theory (math.ST)**; Quantitative Methods (q-bio.QM)

Cite as: **arXiv:1107.2798 [math.ST]**

(or **arXiv:1107.2798v1 [math.ST]** for this version)

Submission history

Which authors of this paper are endorsers?

Link back to: arXiv, form interface, contact.