返回首页

Profile conditional random fields for modeling protein families with structural information

A statistical model of protein families, called profile conditional random fields (CRFs), is proposed. This model may be regarded as an integration of the profile hidden Markov model (HMM) and the Finkelstein-Reva (FR) theory of protein folding. While the model structure of the profile CRF is almost identical to the profile HMM, it can incorporate arbitrary correlations in the sequences to be aligned to the model. In addition, like in the FR theory, the profile CRF can incorporate long-range pairwise interactions between model states via mean-field-like approximations. We give the detailed formulation of the model, self-consistent approximations for treating long-range interactions, and algorithms for computing partition functions and marginal probabilities. We also outline the methods for the global optimization of model parameters as well as a Bayesian framework for parameter learning and selection of optimal alignments.

存档文本