

GraphGibbs Algorithm — A Gibbs Sampling Method for Motif Finding in DNA with Initial Graph Representation of Sequences

Živa Stepančič

Jožef Stefan Institute, Ljubljana

A prominent problem in genetics and bioinformatics is finding short patterns with residue variations, also known as motifs, in DNA sequences. Various motif finding techniques for DNA and protein sequences are inconclusive on real data sets and their performance varies on sequences of different species. In this talk we present an approach to search for possible motifs in DNA sequences in connection to Gibbs sampling method. Starting points in the search space are partly determined via graphical representation of coded initial sequences opposed to completely random choice of initial motif alignment with the standard Gibbs sampling. In addition, information from graphical representation helps us determine the distribution of motif sites per sequence. The performance of our algorithm was evaluated on generated and real data sets using several statistics, such as sensitivity, positive predictive value, specificity, performance and correlation coefficient.