Gibbs Info

Christian Otto

The gibbs sampler approach begins by randomly selecting a motif of length k in each DNA sequence. We then randomly choose one DNA sequence and remove the motif from that sequence from our collection of motifs. We then calculate the profile matrix based on these remaining motifs and calculate the probability for each k-mer occurring in the removed DNA sequence based on the profile-matrix generated. We then select a k-mer by sampling from the distribution of possible k-mers in the DNA sequence that was removed and add it to our collection of motifs. We compare the score of this new collection of motifs to our current best collection, if the score is lower, we select our new collection as the current best. We then repeat the process of randomly selecting a DNA sequence and removing the motif associated with it. We recalculate the profile-matrix and again choose a k-mer based on the probability distribution for all k-mers in the removed DNA sequence. We repeat this process 100 times to give enough time to converge to an optima. We do this entire process starting with a random collection of motifs 1000 times, keeping the collection of motifs with the lowest score as the global best. (Lawrence et al. 1993)

References

Lawrence, Charles E., Stephen F. Altschul, Mark S. Boguski, Jun S. Liu, Andrew F. Neuwald, and John C. Wootton. 1993. "Detecting Subtle Sequence Signals: A Gibbs Sampling Strategy for Multiple Alignment." *Science* 262 (5131): 208–14. https://doi.org/10.1126/science.8211139.