Multi-task Bayesian Model for DNA Motif Search

Hisaki Ikebata
Institute of Statistical Mathematics, Japan

Abstract

A conventional MCMC sampler tends to get stuck in a local mode of a multimodal distribution, and encounters difficulty in escaping to different modes within a finite time of simulation runs. To reach to as many as existing modes, the sampling is often repeated several times with different initial states. Even in doing so, however, a problem would remain; the different chains tend to move and to be trapped to one particular mode having a much higher probability mass. To overcome the difficulties, we propose a new parallel MCMC algorithm. The idea is rather simple: During a parallel run of several MCMC simulations, a repulsive effect is added on each pair of the trajectories. Then the different samplers work in to explore different regions, i.e. a task allocation. With just one-time parallel computation, we would be able to capture as diverse as many modes. After describing the methodology and a link to the parallel tempering algorithm, an application to a pattern mining problem for DNA sequences, called the motif finding problem, will be demonstrated.