

New Kernel Methods for Phenotype Prediction from Genotype data

Ritsuko Onuki^{1,2}, Tetsuo Shibuya¹, Minoru Kanehisa^{1,2}

¹Bioinformatics Center, Institute for Chemical Research, Kyoto University

²Human Genome Center, Institute of Medical Science, University of Tokyo

Phenotype prediction from genotype data is one of the most important issues in computational genetics. In this work, we propose a new kernel (i.e., an SVM: Support Vector Machine) method for phenotype prediction from genotype data. In our method, we first infer multiple suboptimal haplotype candidates from each genotype by using the HMM (Hidden Markov Model), and the kernel matrix is computed based on the predicted haplotype candidates and their emission probabilities from the HMM. We validated the performance of our method through experiments on several datasets: One is an artificially constructed dataset via a program GeneArtisan, another is a real dataset of the NAT2 gene from the international HapMap project, and a real dataset of genotypes of diseased individuals. The experiments show that our method is superior to ordinary naive kernel methods (i.e., not based on haplotype prediction), especially in cases of strong LD (linkage disequilibrium).