

SPADIS: An Algorithm for Selecting Predictive and Diverse SNPs in GWAS

by

Öznur Taştan

Date and Time: May 8th, 2018 (Tuesday), 13:00

Place: Amphi-1, Electronics Engineering Building, GTU

ABSTRACT:

Phenotypic heritability of complex traits and diseases is seldom explained by individual genetic variants identified in genome-wide association studies (GWAS). Many methods have been developed to select a subset of variant loci, which are associated with or predictive of the phenotype. Selecting SNPs that are close on a biological network such as SNP-SNP networks have been proven successful in finding biologically interpretable and predictive SNPs. However, we argue that the closeness constraint favors selecting redundant features that affect similar biological processes and therefore does not necessarily yield better predictive performance. In this work, we propose a novel method called SPADIS that selects a set of loci such that diverse regions in the underlying SNP-SNP network are covered. Instead of enforcing selections based on closeness in the network, SPADIS favors the selection of remotely located SNPs in order to account for the complementary additive effects of SNPs that are associated with the phenotype. This is achieved by maximizing a submodular set function with a greedy algorithm that ensures a constant factor $(1 - 1/e)$ approximation to the optimal solution. We compare SPADIS to the state-of-the-art method SConES, on a dataset of Arabidopsis Thaliana genotype and continuous flowering time phenotypes. SPADIS has better regression performance in 12 out of 17 phenotypes on average, it identifies more candidate genes and runs faster. We also investigate the use of Hi-C data to construct SNP-SNP network in the context of SNP selection problem for the first time, which yields slight improvements in regression performance.

BIOGRAPHY:

Dr. Taştan holds a BSc in Biological Sciences and Bioengineering from Sabanci University and received her PhD in 2011, from Carnegie Mellon University, School of Computer Science. Before joining to Bilkent, she worked as a post-doctoral researcher at Microsoft Research New England Lab (Cambridge, MA, USA). Between 2012-2017, she was a faculty member of the Department of Computer Engineering of Bilkent University. She recently joined Sabanci University as a faculty member of Computer Engineering and Science and Molecular Biology, Genetics Programs and Bioengineering programs. She has worked on diverse problems in computational biology; her present efforts center on building predictive models to advance the current understanding of cancer and other complex diseases.