

Detection of interactions among many potential predictors

Jan Mielniczuk
Institute of Computer Science PAS, Warsaw
miel@ipipan.waw.pl

1. Project Description

The aim of the project is to propose new methods, and compare them with the existing ones, of detecting epistatic interactions in Genome-Wide Association Study (GWAS) data, which examines associations between genotypes and phenotypes in a study group. We will focus on interactions between Single Nucleotide Polymorphisms (SNPs) in explaining binary trait. There are three, partly unresolved issues, which we will try to address: how to define epistatic interaction and its contextual dependence, how to measure it and how to avoid overwhelming computational complexity of exhaustive search among all possible pairs of SNPs. For the second problem, we will focus on Interaction Information measure which is a model free method which frequently detects interesting interactions which remain undetected by model-based methods. However, there are still interesting theoretical and applicational problems pertaining to it, such as determination of its distribution when only main effects of two SNPs are present. The third problem of finding computationally feasible search method is of paramount importance. The required screening-selection solution should retain after screening step pairs of genes without negligible main effects but with significant interaction. The sought-after method will be tested against known popular algorithms such as TEAM, BOOST, SNPHarvester and Screen and Clean among others. For review of the problem consult Upton et al, Review: High-performance computing to detect epistasis in genome scale data sets, Briefings in Bioinformatics 2015.

2. Requirements

- a. M.Sc. of Computer Science, Mathematics or Physics
- b. Programming skills
- c. Elementary knowledge of machine learning, statistics, optimisation and numerical methods
- d. Eagerness to learn and to solve problems
- e. Good command of English