A walk into random forests. Adaptation and application to Genome-Wide Association Studies

Vincent Botta (ULg)
November 13, 2013

The primary goal of Genome-Wide Association Studies (GWAS) is to discover genetic markers associated with a disease. Standard approaches to GWAS are usually based on univariate hypothesis tests and therefore cannot account neither for linkage disequilibrium nor for combinations of several markers. This thesis studies the application of supervised machine learning and, in particular, tree-based methods in the GWAS context. This investigation leads to an extension of the Random Forest algorithm tailored for structured GWAS data and its systematic validation on 7 different real-life datasets.

Webpage: http://www.montefiore.ulg.ac.be/~botta/