
Classifying with Decision Diagrams Obtained From Genetic Programming

Robin Nunkesser

Faculty of Computer Science, TU Dortmund University
Robin.Nunkesser@tu-dortmund.de

Abstract. Our main focus is on genetic association studies concerned with single nucleotide polymorphism (SNP) data, i.e. data on genetic variations that occur when different base alternatives exist at a single base pair position. Typical classification on this data aims at distinguishing between two groups, e.g. diseased subjects (cases) and healthy controls. Genetic Programming has already been successfully applied to case-control genetic association studies. However, some applications need to distinguish between more groups and many promising methods for genetic association studies cannot originally handle more than two groups.

We present a Genetic Programming algorithm that allows to distinguish between two and more groups. The basis to do this is to use Decision Diagrams as prediction models. Decision Diagrams are originally a data structure for Boolean functions that may be seen as a special classification tree. Using Decision Diagrams instead of more general decision trees allows faster operations in the algorithms used. The algorithm is implemented in an extendible R package called RFreak, which also allows an easy modular implementation of further Genetic Programming and Evolutionary Computation algorithms.

Additionally, we demonstrate the good performance of the algorithm on simulated and real data.

Key words: Association Studies, Decision Diagrams, Genetic Programming, Multicriteria Optimization

References

- 1.J.R. Koza. *Genetic Programming*. The MIT Press, Cambridge, Mass., 1992.
- 2.R. Nunkesser, T. Bernholt, H. Schwender, K. Ickstadt, and I. Wegener. Detecting high-order interactions of single nucleotide polymorphisms using genetic programming. *Bioinformatics*, 23(24):3280–3288, 2007.
- 3.I. Wegener. *Branching Programs and Binary Decision Diagrams*. SIAM, Philadelphia, 2000.