Inferring gene regulatory networks using ensembles of feature selection techniques

Joeri Ruyssinck[1], Vân anh Huynh-thu[2], Pierre Geurts[2], Tom Dhaene[*], Piet Demeester[1], Yvan Saeys[3]

[1] Department of Information Technology, Ghent University – IBBT, Gent, Belgium, [2] Department of Electrical Engineering and Computer Science, Systems and Modeling, University of Liège, Liège, Belgium, [3] Department of Plant Systems Biology VIB, Ghent University, Gent, Belgium

Motivation: One of the long-standing open challenges in computational systems biology is the inference of gene regulatory networks from expression data. Recently, two community-wide efforts have been established to benchmark network inference techniques (DREAM4 and DREAM5), where it was shown that a feature selection method us- ing variable importance scores obtained from tree-based ensemble methods (GENIE3) achieved top performance. Despite the success of this algorithm, little research has been carried out to understand why this approach works so well, and if equally good or better results could be obtained using other types of feature selection techniques.

Results: In this work, we present a large scale analysis of feature selection approaches to the network inference problem. We show that, using the recent concept of ensemble feature selection techniques, equally good or better results than GENIE3 can be achieved, demonstrating that the ensemble setting is a necessary requirement for feature selection techniques to achieve good performance on the network inference task. Furthermore, we show that by combining several ensemble feature selection techniques the performance can be made more robust and slightly improved. This analysis opens up new avenues for the development of novel types of ensemble based feature selection techniques in this setting.

Contact: joeri.ruyssinck@ugent.be