Fundamenta Informaticae 113 (2011) i-ii

DOI 10.3233/FI-2011-601

IOS Press

Machine Learning in Bioinformatics

Preface

Computational biology is an application domain where information is naturally represented in terms of relations between heterogeneous objects. Modern experimentation and data acquisition techniques allow the study of complex interactions in biological systems. This raises interesting challenges for machine learning and data mining researchers, as the amount of data is huge, some information cannot be observed, and measurements may be noisy.

This special issue collects three papers on machine learning methods designed to deal with the specific challenges posed by the field of computational biology. Some of these papers are extensions of contributions at the two workshops on Statistical and Relational Learning in Bioinformatics at ECML 2008 and KDD 2009. These were organized to bring together researchers of the machine learning field with specific interest in developing statistical and relational approaches to bioinformatics.

The main problems analyzed in this issue include relation induction, and knowledge sharing between related tasks. Modeling biological problems in a relational framework offers the advantage of a richer representation for expressing dependencies among heterogeneous entities. However the reliable induction and assessment of these dependencies in complex computational biology domains is a challenging task. In the first paper, J.M. Arevalillo and H. Navarro present a method, based on random forests, which is particularly robust and capable to detect pairwise dependencies in high dimensional data. The method is then applied to gene expression data. In the second paper, W. Hämäläinen focuses on discovering dependencies amongst multiple entities in an efficient way. The dependency rules are mined according to typical statistical measures, and redundant rules are pruned in order to avoid blurring of dependencies and erroneous conclusions.

Re-using and sharing knowledge from related tasks can improve the performance of predictors. An effective exploitation of this concept in relational models is however a challenging problem. The last paper, by E. Cilia, N. Landwehr and A. Passerini, introduces a hierarchical kFoil algorithm. This algorithm is designed to learn multiple targets simultaneously and to then further specialize for each target or target cluster separately. The paper then applies this approach to several bioinformatics problems to demonstrate that the algorithm is both useful and boosts performance.

We thank the reviewers of the papers in this special issue for their valuable contributions and the authors for their careful consideration of the reviewers' comments. This resulted in a good quality of the included papers.

Editors

Jan Ramon
Katholieke Universiteit Leuven, Belgium

Fabrizio Costa
Katholieke Universiteit Leuven, Belgium

Christophe Costa Florêncio
University of Amsterdam, Netherlands

Joost Kok
Leiden University, Netherlands