Preface

The ECML Workshop on Statistical and Relational Learning in Bioinformatics (StReBio) will be held at the University of Antwerp, Antwerp, Belgium, on September 19, 2008. This workshop is intended to bring together researchers from both the fields of statistical and relational learning and the field of bioinformatics. Hopefully this will contribute to communication between these communities and will stimulate the development of new relational techniques for application in the biological domain.

Bioinformatics is a domain where information is naturally represented in terms of relations between heterogenous objects. Modern experimentation- and data acquisition techniques make the study of complex interactions in biological systems possible, yielding huge amounts of data. This raises challenges for the machine learning and data mining communities, where interest in relational and statistical learning has been growing in the last few years. Apart from the amount of (relational) data, the information is often incomplete, and measurements may be noisy, creating additional challenges.

We have tried to choose a workshop format which promotes as much as possible the exchange of ideas. Aspects which can help to achieve our goal are the problem statement presentations, the time allocated in the program for questions and discussion, the informal proceedings and the workshop website (http://www.ecmlpkdd2008.org/StReBio) where the abstracts, papers and some presentations will be available.

We selected four papers for full presentation and three for poster presentation. All these contributions, in varying degrees, fit the format of the workshop: they apply statistical or relational techniques to problems from the biological domain. Furthermore, we accepted three problem statements to be presented at the workshop.

The idea of the problem statement track is to stimulate the interaction between biologists and computer scientists by presenting biological problems in a more mathematical form, making more explicit structure and uncertainty of the data. We hope that the problem statements will lead to lively discussions and the exchange of ideas. We feel honoured to have Stephen Muggleton and Walter Luyten as invited speakers, who, in the spirit of this workshop, are authorities in the fields of machine learning and bioinformatics, respectively.

The organizers would like to acknowledge the (financial) support from the Machine Learning group of the Department of Computer Science at K.U. Leuven and the ECML/PKDD organisation. We like to thank the reviewers and program committee members for their support. Furthermore, we thank ECML/PKDD's Organizing Comittee for their help and making this workshop possible.

Christophe Costa Florêncio, Jan Ramon, Fabrizio Costa and Joost Kok
August 2008

Program committee

Andreas Bender (Leiden/Amsterdam Center for Drug Research, The Netherlands)
Amanda Clare (University of Aberystwyth, UK)
Luc Dehaspe (Oncomethylome Sciences, Belgium)
Paolo Frasconi (University of Florence, Florence, Italy)
Yves Moreau (ESAT, K.U.Leuven, Belgium)
Marcel Reinders (TU Delft, The Netherlands)
Herman Van Vlijmen (Tibotec, Belgium)

Organizing committee

Jan Ramon, K.U. Leuven, Belgium
Fabrizio Costa, K.U. Leuven, Belgium
Christophe Costa Florêncio, K.U. Leuven, Belgium
Joost Kok, Leiden Institute of Advanced Computer Science, The Netherlands