

Introduction

Knowledge discovery in bioinformatics

José-María Peña^a and Evgenii Vityaev^b

^a*Universidad Politécnica de Madrid, Spain*

^b*Sobolev Institute of Mathematics, Russia*

We'd like to welcome you to this special issue of the Intelligent Data Analysis journal devoted to Knowledge Discovery in Bioinformatics. The articles selected for this issue partially come from the special session with the same name, Knowledge Discovery in Bioinformatics, in BGRS 2008. Additional articles, on the same topic, have been included from regular contributions received by the IDA journal.

The special session on Knowledge Discovery in Bioinformatics was the third of this kind of successful sessions hosted by three different conferences worldwide, the European Conference in Artificial Intelligence – ECAI 2004 (Valencia, Spain), the European Conference of Machine Learning – ECML/PKDD 2007 (Warsaw, Poland) and, finally, the Sixth International Conference on Bioinformatics of Genome Regulation and Structure – BGRS 2008 (Novosibirsk, Russia). The aim of these sessions has always been to show the interdisciplinary research carried out in the joint collaboration of data analysis/modeling and system biology, genomics or proteomics.

The process of selecting the papers among the contributions submitted for the BGRS special session was a tough task. There was a great quality in the works presented and a broad variety in terms of both techniques and applications. Finally, those best representing the on-going research shown in the session were asked for an extended version including new results and a more detailed description of the techniques put in practices. The author had the opportunity to extend their work for few more months before a peer-reviewed process for the selection of the first four articles included in this selection:

The article *PROMETHEUS: Technology for Rapid Development of Biological Databases* by Timonov and Miginsky presents a new technology for the development of biological databases, specifically designed for researchers and practitioners in different life sciences domains.

Beslon et al., in their article entitled *From Digital Genetics to Knowledge Discovery:*

Perspectives in Genetic Network Understanding, present an integrated approach to model regulatory networks in artificial organisms. The proposed model is compared by means of a data analysis process, similar to the one performed on living organisms (effect of gene knock-out in metabolic networks), including discussion about the application of artificial life methods as a benchmark for the study of biological networks.

In the article *Independent Component Analysis for Microarray Data Analysis* by Malutan, Gómez-Vilda and Borda, the authors present a new methodology for the detection process in microarrays by multichannel differential expression. This new methodology, based on genomic signal processing, is able to identify unexpected hybridization dynamics.

The article *CliDaPa: A New Approach to Combining Clinical Data with DNA Microarrays* by González et al. introduces a new algorithm to combine clinical data with different genomic/proteomic data. The

proposed method uses clinical data to perform data partitioning on the set of patients, which are classified using the gene expression data with specific models for each data partition.

To make this special issue a full publication, two additional papers that are closely related to the topic of life sciences and were accepted for publication have been added to this issue. Parthalan et al in the article entitled: Fuzzy-rough approaches for mammographic risk analysis present a unified approach for the application of a number of rough and fuzzy set methods to the analysis of mamographic data for breast cancer patients. The approach is based on using feature selection methods to reduce the data dimensionality and applying a number of classifiers to examine reduced data and asses the positive impact of data reduction on classification accuracy. Their approach is evaluated using two well known data sets. Chen et al in the last article of this issue entitled: Mining consequence events in temporal health data formalize the problem of mining consequence events and combine the idea of Before-After-Control-Impact design with frequent pattern mining techniques to define an interestingness measure called consequence. The idea is evaluated on some real life science data where consequence events result in lowest and highest consequence values for negative and positive cases, respectively.

We would like to thank Fazel Famili for encouraging us to prepare this selection of articles, giving us the possibility to publish them in the IDA journal. We would like to mention the support offered also by the BGRS organizers, in particular Nikolay Kolchanov, Ralf Hofestadt, Dagmara Furman, and Ilya Akberdin. Finally, we must thank also the external reviewers that kindly provided valuable comments and suggestions in the selection of the final papers: Stan Matwin, Min-Ling Zhang, Wolfgang Banzhaf, Jorge Amigo, Delphie Ropers, Ana Tome and Borja Calvo.