Guest Editorial

Welcome to this special issue of *Intelligent Data Analysis*.

This special issue is devoted to the new methods in bioinformatics presented at one of the most representative conference on bioinformatics in Russia: "Fifth International Conference on Bioinformatics of Genome Regulation and Structure" – BGRS'2006 (Akademgorodok, Novosibirsk, Russia, July 16–22, 2006 http://www.bionet.nsc.ru/meeting/bgrs2006/).

Nowadays, biology is among the most rapidly developing scientific disciplines. The main factor of this progress is unprecedented, both in rate and volume, accumulation of new facts due to advent of novel state-of-the-art experimental technologies. The post-genome era in biology brought about intensification of research in the fields of genomics, transcriptomics, and proteomics. We witness the emergence and advance of new directions in experimental and computer molecular biology, including sequencing and analysis of megagenomes of bacterial communities, regulation of gene expression by short RNAs, microarray analysis technique, construction of proteomic portraits of cells and tissues, metabolomics, highly efficient genotyping of human populations for biomedical purposes, and many others.

However, the synthesis of these directions lags behind, while it is a primary need for creation of an orderly theory of development, function, and evolution of living systems: systems biology (gene interaction, gene network functioning, signal transduction pathways, networks of protein-protein interactions, modeling of ontogenesis, molecular phylogeny, theory of evolution, etc.). The reasons underlying this gap lie not only in the complexity of living systems but also in the specialization in various fields of biology, which is ever increasing with accumulation of new data and development of new methods. The holistic vision of the research object is disappearing.

The goal of BGRS'2006 was to provide an opportunity for wide exchange of opinions between experts *in silico* biology and researchers involved in experimental studies who use computer methods in their work or have interest in applied or theoretical aspects of bioinformatics. The conference provided a general forum for disseminating the latest developments in bioinformatics in molecular biology.

The holistic vision of Machine Learning and KDD&DM application in bioinformatics is also disappearing. Furthermore, the current state of the methods does not correspond to the scope, scale and diversity of the bioinformatics tasks. Bioinformatics tasks are challenges to the Machine Learning and KDD&DM methods:

– huge volume and diversity of available data, requiring methods of processing heterogeneous databases;
– heterogeneity of bioinformatics processes (discrete and continues, deterministic and stochastic, precise and fuzzy), requiring hybrid methods;
– hierarchical organization of bioinformatics processes, requiring methods of hierarchical data analysis;
– systemic character of bioinformatics processes, requiring methods operating with large number of interconnecting parameters.

In this volume we gathered the papers that concern the bioinformatics tasks challenges.

The volume and diversity of the available data is described in the paper of Kolchanov N.A. et al. "TRRD: technology for extraction, storage, and use of knowledge about the structural-functional organization of the transcriptional regulatory regions in the eukaryotic genes". In this paper the Transcription Regulatory Regions Database (TRRD, http://www.bionet.nsc.ru/trrd/), designed for storage of data on the structural-functional organization of transcriptional regulatory regions of the eukaryotic genes and their expression patterns, is described.

The heterogeneity, hierarchy and systemic character of the bioinformatics processes are considered in the paper of Miginsky D.S. et al. "Approaches to the Computer Reconstruction of the Biological Networks", where a network modelling technique for effective biological system modelling, analysis and simulating is described. In the paper a unified software toolkit for describing ecological and molecular-genetic systems of high complexity (of thousands and even more elements) is presented.

Hierarchical data analysis is presented in the paper by Khomicheva I.V. et al "ExpertDiscovery system application for the hierarchical analysis of eukaryotic transcription regulatory regions based on DNA codes of transcription". In this paper the special relational data mining approach and program system 'Discovery' is adapted to analysis of eukaryotic transcription regulatory regions. As a result, the 'ExpertDiscovery' system is developed, that discovers a hierarchically complicating set of complex signals based on different elementary signals.

A hierarchical classification is presented in the paper by Vityaev E.E. et al. "Transcription factor binding site recognition by regularity matrices based on the natural classification method". A principally new approach to the classifications tasks is presented based on the concept of "natural" classification. The "natural" classification is hierarchical by the nature. As a result of "natural" classification of the nucleotide sequences regularity matrices are produced, where nucleotides are interconnected by regularities.

An example of hybrid methods is presented in Victor Levitsky's paper "Genetic algorithm and optimized weight matrix application for peroxisome proliferator response elements recognition: prerequisites of accuracy growth for wide genome research". The SiteGA model is proposed that uses a combination of the genetic algorithm and a discriminant function of locally positioned dinucleotides for prediction of the most important positions and dinucleotides for binding sites (BSs) of peroxisome proliferator-activated receptor.

The presented papers outstep the standard application of the Machine Learning and KDD&DM methods to bioinformatics tasks and they are only first steps in the analysis of bioinformatics processes.

Guest Editors
Evgenii Vityaev and Nikolai Kolchanov