

## RECENT ADVANCES IN OPERATIONS RESEARCH IN COMPUTATIONAL BIOLOGY, BIOINFORMATICS AND MEDICINE

METIN TURKAY<sup>1</sup>, GIOVANNI FELICI<sup>2</sup>, MARTA SZACHNIUK<sup>3</sup> AND PIOTR LUKASIAK<sup>4</sup>

**Abstract.** The EURO Working Group on Operations Research in Computational Biology, Bioinformatics and Medicine held its fourth conference in Poznan-Biedrusko, Poland, June 26–28, 2014. The editorial board of RAIRO-OR invited submissions of papers to a special issue on Recent Advances in Operations Research in Computational Biology, Bioinformatics and Medicine. This special issue includes nine papers that were selected among forty presentations and included in this special issue after two rounds of reviewing.

**Mathematics Subject Classification.** 92Cxx.

Received September 8, 2015. Accepted September 21, 2015.

### 1. INTRODUCTION

The EURO Working Group on Operations Research in Computational Biology, Bioinformatics and Medicine (CBBM) was established in 2006 to foster collaboration in the development of models, solutions algorithms and data analysis approaches to emerging problems in life sciences. Our ability to generate data related to biological systems from molecular level to physiological level increased dramatically over the past decade, thanks to innovative developments in the experimental apparatus. However, the analysis of data and also interpretation of the results to develop insights into complex biological system is still one of the most important bottlenecks for the further advancement of life sciences. Operations research already has established roots in the sectors of biology, health and medicine and, through its quantitative and interdisciplinary instruments, is able to enrich the interpretation of the results through data analytics, modeling, simulation and optimization. The EURO Working Group CBBM organized four conferences after its establishment:

- I WG CBBM Conference, Prague, Czech Republic, 8 July 2007,
- II WG CBBM Conference, Rome, Italy, 15–17 September 2008,
- III WG CBBM Conference, Nottingham, United Kingdom, 13–15 September 2012,
- IV WG CBBM Conference, Poznan-Biedrusko, Poland, 26–28 June, 2014.

---

*Keywords.* Operations research, computational biology, bioinformatics.

<sup>1</sup> Department of Industrial Engineering, Koc University, 34450 Istanbul, Turkey. [mturkay@ku.edu.tr](mailto:mturkay@ku.edu.tr)

<sup>2</sup> Consiglio Nazionale delle Ricerche, Istituto di Analisi dei Sistemi ed Informatica, A. Rubertiâ, Roma, Italy. [giovanni.felici@iasi.cnr.it](mailto:giovanni.felici@iasi.cnr.it)

<sup>3</sup> Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland. [mszachniuk@cs.put.poznan.pl](mailto:mszachniuk@cs.put.poznan.pl)

<sup>4</sup> Institute of Computing Sciences, Poznan University of Technology, Poznan, Poland. [plukasiak@cs.put.poznan.pl](mailto:plukasiak@cs.put.poznan.pl)

The EURO Working Group CBBM was involved in the preparation of three special issues in the past:

- The first special issue was dedicated to the networks in computational biology [4]. This special issue was focused on the models and algorithms for the analysis of biological networks.
- The second special issue was prepared among the papers that were presented in the II WG CBBM Conference in Rome [1]. It included papers mainly focusing on models and algorithms for analysis of biological data.
- The third special issue was prepared in collaboration with EURO Working Group OR Applied to Health Care among the papers that were presented in the 25th European Conference on Operational Research in Vilnius, Lithuania [12]. The papers included in this special issue were diverse, covering data analysis, disease diagnostics, models and algorithms for the analysis of biological systems.

With this special issue, we summarize the state-of-the-art contributions presented and discussed in the occasion of IV WG CBBM Conference in Poznan-Biedrusko, Poland. Our motivation is to further encourage excellent research and interaction between OR and life sciences.

## 2. PAPER SUMMARIES

This special issue of RAIRO-OR on the Recent Advances in Operations Research in Computational Biology, Bioinformatics and Medicine consists of 9 articles, written by 29 authors and rigorously reviewed by 35 referees. The guest editors express their cordial thanks to all of them, as well as to the Editor-in-Chief and editorial board members of RAIRO-OR.

The articles are listed according to their contents. Let's examine them briefly.

The research on the analysis of DNA sequences was a pivotal problem in the development bioinformatics. The paper by Lancia [8] presents the Single Individual Haplotyping problem. The paper argues that one should find and correct the errors in the data as well as retrieve a consistent pair of haplotypes when inconsistent haplotype data comes from the sequencing of an individual's chromosome. An overview of several methods and models used to solve the problem are discussed and analyzed, presenting wide range of operational research approaches.

The determination of three-dimensional structure of proteins and protein complexes is an important problem. The paper by Cetinkaya *et al.* [2] addresses the backbone assignment problem appearing in the process of NMR-based determination of three-dimensional structure of proteins. The authors have developed a new method for protein assignment evaluation. A contribution to speeding up of the NVR approach and extending its capability to handle various types of data, and to process noised spectra, has made the method more practical and useful in laboratory work.

While the collection and analysis of biomedical and genomic data are well-recognized challenges, a third challenge is also very important: the efficient storage and maintenance of large collection of genomic data (in particular, data extracted with new generation sequencing technologies). This challenge is discussed in the paper by Chlopkowski *et al.* [3], where a very efficient algorithm for compressing FASTQ files produced by high-throughput DNA sequencers is presented. The authors describe the algorithm in detail, summarize experimental results, and show how their FastQComp algorithm achieves better compression ratio than other compressors specialized for FASTQ files.

Milostan [9] presents a new graph based method that can be used for recognition of domains – conserved parts of protein structure that are associated with particular functions of proteins. Proposed solution uses a specially crafted contact graph and graph clustering to detect potentially stable substructures, and has been tested on the data from SCOP and CATH databases. The authors show that the method can be easily applied for the considered problem and extended by additional analytical modules devoted to recognition of similarity of domain cores in proteins.

The paper by Kulik *et al.* [7] investigates the hydrogen bond patterns, solvent accessibility, and stacking energies of the nucleobases in twelve different aminoglycoside-RNA crystal complexes. Their analyses show some antibiotic-induced RNA structural differences in the hydrogen bonding pattern in the vicinity of the

U1406 and U1495 base pairs, especially in the complexes with geneticin and modified paromomycin. They also investigated the dependence of other aminoglycoside characterizing descriptors, such as the number of rings and total charge, on the experimentally determined Gibbs energies.

The structure-based drug design takes the three-dimensional structure of the target protein and tries to find molecular structures that are able to regulate the functions of the target protein through molecular interactions. Tardu *et al.* [11] applies hyperbox classification method to determine the most relevant molecular descriptors of drug molecules that inhibit the activity of Sirtuin6. The comparison of the proposed approach with the other approaches has shown that the the hyperbox approach performs very well on the classification of activities of Sirtuin6 inhibitors.

The analysis of molecular dynamics (MD) is a difficult problem due to the fact that there are many alternative trajectories and finding the most plausible trajectory is an unsolved problem. Gogolinska *et al.* [5] presents a number of algorithms for generating Petri Nets (PN) from a large set of MD trajectories. Among the algorithms, One Place One Conformation (OPOC), is presented in a greater detail that reflects the changes between biomolecule conformations. In addition, three simulations are analyzed to show the effectiveness of the OPOC algorithm. The authors show that PNs reproduce events hidden in MD trajectories and enable observations of the conformational space features hard-to-see by the other clustering methods.

Kropat *et al.* [6] presents the concept of fuzzy target-environment networks together with various fuzzy possibilistic regression models, for the analysis of two-modal regulatory systems affected by errors and uncertainty. The analyzed fuzzy regression approaches are very flexible and can be adapted to a variety of regulatory systems where data uncertainty and model restrictions are involved. This is usually the case for interdependent networks in systems biology, where the regulating effects are not known in detail and one has to refer to approximating models for a representation of the interaction at the system level.

The paper by Santoni *et al.* [10] describes a new method to measure the contribution of discretized features for supervised learning, presenting its usefulness for biological data analysis. The authors show the ability of the proposed approach to score and rank features for supervised learning in a more effective way than others well established rankers. The experimental results highlight the potential advantages of using this method in the analysis of biological data sets, where few relevant continuous features must be extracted from large sets, and properly discretized.

### 3. CONCLUSIONS

This special issue of RAIRO-OR on the Recent Advances in Operations Research in Computational Biology, Bioinformatics and Medicine includes only a limited sample of papers on a variety of subjects on computational, bioinformatics and medicine. The editors of the special issue share the vision that the operations research methods will more and more play a significant role in the analysis of biological data, and in the modeling, simulation and optimization of complex biological systems.

### REFERENCES

- [1] P. Bertolazzi, J. Blazewicz and M. Turkey, Operations research models for computational biology, bioinformatics and medicine. *J. Math. Model. Algorithms* **9** (2010) 209–211.
- [2] S. Cetinkaya, S.N. Ekren and M.S. Apaydin, Progress in Nuclear Vector Replacement For NMR Protein Structure-Based Assignment. *RAIRO: OR* **50** (2016) 341–349.
- [3] M. Chlopkowski, M. Antczak, M. Slusarczyk, A. Wdowski, M. Zajackowski and M. Kasprzak, High-order statistical compressor for long-term storage of DNA sequencing data. *RAIRO: OR* **50** (2016) 351–361.
- [4] A. Dress, B. Karasozen, P.F. Stadler and G.W. Weber, Networks in computational biology. *Discrete Appl. Math.* **157** (2009) 2217–2220.
- [5] A. Gogolinska, R. Jakubowski and W. Nowak, Petri nets formalism facilitates analysis of complex biomolecular structural data. *RAIRO: OR* **50** (2016) 407–417.
- [6] E. Kropat, A. Ozmen, G.W. Weber, S. Meyer-Nieberg and O. Deftlerli, Fuzzy prediction strategies for gene-environment networks – Fuzzy regression analysis for two-modal regulatory systems. *RAIRO: OR* **50** (2016) 443–455.

- [7] M. Kulik and J. Trylska, Structural and energetic comparison of aminoglycosides with the model of the ribosomal A-site. *RAIRO: OR* **50** (2016) 381–392.
- [8] G. Lancia, Algorithmic approaches for the Single Individual Haplotyping problem. *RAIRO: OR* **50** (2016) 331–340.
- [9] M. Milostan and P. Lukasiak, DOMGEN – Graph based method for protein domain delineation. *RAIRO: OR* **50** (2016) 363–379.
- [10] D. Santoni, E. Weitschek and G. Felici, Optimal discretization and selection of features by association rates of joint distributions. *RAIRO: OR* **50** (2016) 443–455.
- [11] M. Tardu, F. Rahim, I.H. Kavakli and M. Turkay, MILP-hyperbox classification for structure-based drug design in the discovery of small molecule inhibitors of Sirtuin6. *RAIRO: OR* **50** (2016) 393–406.
- [12] G.W. Weber, J. Blazewicz, M. Rauner and M. Turkay, Recent advances in computational biology, bioinformatics, medicine and healthcare by modern OR. *Central Eur. J. Oper. Res.* **22** (2014) 427–430.