Biomedical and Bioinformatics Challenges to Computer Science: Bioinformatics, Modeling of Biomedical Systems and Clinical Applications

Mario Cannataro*, Rodrigo Weber dos Santosb, Joakim Sundnesc

aDepartment of Experimental Medicine and Clinic, University "Magna Græcia" of Catanzaro, Catanzaro, Italy, and ICAR-CNR, Rende, Italy
bDepartment of Computer Science, Universidade Federal de Juiz de Fora, Juiz de Fora, Brazil
cCenter for Biomedical Computing, Simula Research Laboratory, Lysaker, Norway

Abstract

This is the fourth edition of the workshop on Biomedical and Bioinformatics Challenges to Computer Science. The purpose of the workshop series is to bring together scientists from computer science and life sciences, to discuss current challenges in this inter-disciplinary field. A wide range of life science applications will be addressed, ranging from classical bioinformatics to mathematical models of systems physiology.

Keywords: Bioinformatics, modeling and simulation of biological systems, biomedical data management and integration, clinical applications

1. Introduction

Computers and computational methods play an increasingly important role in the life sciences, and currently form an indispensable tool for a wide range of scientific applications. These range from classical bioinformatics, which concerns with the use of computer methods in molecular biology, to numerical models of complex physiological systems. To this date, these scientific fields have been fairly disconnected, but the situation is changing with the increasing availability of high performance computing tools, which allows system physiology models to be developed with more detail and realism. Existing computer models have already demonstrated direct links between gene data and organ function, and the current development will inevitably narrow the gap between classical bioinformatics and systems physiology modeling. The successful combination of these two fields holds a huge potential for scientific progress, as it may represent a mechanistic link from knowledge and exploration on a molecular scale, to the function of biological tissues, organs and organ systems. However, this large potential is paired with substantial challenges, many of which are directly related to computer science. These include the management and integration of huge data banks, computational resources required for solving complex multiscale mathematical models, and software engineering issues related to development and validation of increasingly complicated software systems.

*Corresponding author: cannataro@unicz.it
The purpose of the fourth workshop on Biomedical and Bioinformatics Challenges to Computer Science is to bring together scientists from computer science, applied mathematics and a variety of branches of life sciences, to discuss important computer science challenges.

2. Workshop summary

The papers of the present workshop cover a broad range of applications and solutions in Bioinformatics, Modeling of Biomedical Systems and Clinical Applications.

The first session contains four papers devoted to Bioinformatics. Three papers are devoted to modeling and analysis of genomics data: one paper describes novel approaches to solve applications of strings that are useful in genomics studies; one paper discusses methods to improve the prediction of protein interaction from gene expression data; and one paper presents a stochastic model for the micro RNA (miRNA) - messenger RNA (mRNA) network. The last paper of the session describes how bioinformatics services can be modelled and queried using the novel Search Computing paradigm.

The paper by Kelsey and Kotthoff proposes the Constraint Satisfaction paradigm as a computational framework for solving the closest string problem. A Constraint Satisfaction Problem specifies a set of constraints on discrete values variables and a solution to such a problem is an assignment of values to variables such that no constraint is violated. Given a set of strings of equal length over a fixed alphabet, the closest string problem consists on finding the string with the smallest possible maximum Hamming distance from any input string. When the input strings consist of nucleotide sequences over the letters A, C, G and T, the closest string has important applications in computational biology. The authors propose search heuristics based on Constraint Satisfaction that provide several orders of magnitude speed-up at and above the optimal distance.

The paper by Yano evaluates different methods to measure correlation among gene expression levels. In microarray data analysis, measuring correlation among gene expression levels is a common practice to predict interactions among gene products. Correlated expression of a group of genes implies that the genes are involved in the same biological process or form a protein complex, and correlation measures have been used to predict disease markers and protein-interaction partners. The paper evaluates the performance (sensitivity and specificity) of three correlation methods: Pearson correlation, rank (Spearman) correlation, and Mutual Information, for the detection of protein-protein interactions, and proposes a new correlation measure, termed asymmetric correlation (AC), which outperforms the other measures, particularly when high specificity is required.

The paper by Arpag et al proposes a stochastic model for the micro RNA (miRNA) - messenger RNA (mRNA) network. miRNAs are 20-22 nucleotide long strands produced from non-coding RNA, and interfere with the translation of mRNA by binding their untranslated regions (UTR) at the 3’ end of mature mRNA. While existing computational methods predict target sequences that are likely to be bound by particular miRNA, the paper investigates the overall statistical properties of the mRNA-miRNA network under the assumption that all the sequences are stochastic. The obtained analytical and simulation results for the distribution of the number of miRNA regulating any given mRNA, compare very well with the empirical data. Authors conclude that a major portion of the miRNA-mRNA interactions can be accounted for purely on the basis of combinatorics, even though some elements of this regulatory mechanism are highly conserved.

The paper by Masseroli et al addresses the problem of answering complex biomedical questions that require the querying and integration of different distributed bioinformatics services, that frequently are inherently ordered, or are associated with ranked confidence values. To solve such problems they propose to use the Search Computing, an emerging paradigm for integrating and querying web services. Search Computing aims to answer complex questions by combining data extracted from distinct sources, ordered based on local rankings, into result combinations, with an associated global ranking, which represents the solution of the complex query. The authors discusses how bioinformatics web service can be described in the search computing framework and present a bioinformatics search computing (Bio-SeCo) application based on different bioinformatics web service described and registered using the Search Computing paradigm.

The second session contains four papers, which are all devoted to the computational modeling and to computer systems related to medical research. Two papers address multinational infrastructures for integrating, sharing and processing information in cancer research; one paper describes a system for acquisition and automatic analysis of
snoring signals for diagnosing sleep apnoea syndrome (SAS); and another paper describes an adaptive time stepping scheme for models of cardiac cell electrophysiology.

The paper by Bucur et al addresses lessons learned from the ACGT project (www.eu.acgt.org). This project was an EU funded Integrated Project (IP), that was launched in 2006 to develop new and improved infrastructure for combining information from genetic and clinical trials, and utilize the potentially vast amounts of information for improved, individualized cancer therapy. Important conclusions from the project were that although some low-level technical decisions may have been suboptimal or have later been superseded by technological development, the overall methodology and core design principles remain valid today. The ontology driven approach, focus on patient data security, and constant emphasis on simplifying the users’ daily work are critical for the success of similar endeavors in the future.

Sinnott and co-workers also address cancer research, and describe a recently established EU funded project ENS@T-CANCER, which aims to establish a Virtual Research Environment (VRE) for the study of adrenal cancer. The plan is for the VRE to comprise a number of clinical databases for various types of adrenal tumors, and offer advanced services for accessing and integrating information on adrenal tumors across organizational and national borders. The project is currently being established, and many of the goals and methodologies of the project are likely to change on the way. However, the ENS@T-CANCER registry already includes records of 183 patients with different forms of adrenal cancer, many of which include biosamples that can be exchanged.

The paper by Calabrese et al describes a system for analysis of snore signals, which is intended to support medical doctors in the diagnosis of sleep apnoea syndrome (SAS) and subsequent follow-up. The system is based on sound recording, and on segmenting sound signals into snoring and regular breathing. The extracted snoring intervals are analyzed in LabView, to detect and characterize events of sleep apnoea. The existing prototype system enables clinicians to record, collect and analyze snoring signals from possible SAS patients. The prototype includes a simple graphical user interface, to facilitate easy use by clinicians in daily diagnosis and follow-up of SAS.

Campos and co-workers present an adaptive time-stepping scheme for numerical solution of models of cardiac cell electrophysiology. These models are typically formulated as systems of non-linear ODEs, with solutions that typically consist of one very rapid change (depolarization) followed by a more smooth restoration of the normal state (repolarization). These characteristics favor adaptive time stepping schemes, which can capture the rapid variations with small time steps and use longer time steps through the smoother intervals. The authors present adaptive methods based on the Euler method and a second order Runge-Kutta method, and preliminary results indicate speed-up of up to 25 times compared with fixed time step methods.

3. Conclusion and outlook

The topics addressed in this workshop represent a wide range of scientific fields, but the contributions can largely be divided into two groups. These groups can be related to the physical scales considered; classical bioinformatics applications concerning the molecular scale, and systems physiology models operating on the scale of organs and organ systems. Although the physiology models typically link down to the cell level, there is still a gap between this field and the classical bioinformatics field, both in physical scales and in scientific tools and modeling approaches. As noted in the previous editions of the workshop [1, 2, 3], a goal for the present workshop series is to increase the interactions between these two branches of science, potentially leading to extremely promising collaborative research projects.

Acknowledgements

We would like to thank the members of the program committee for the invaluable assistance on reviewing the papers and the organizers of ICCS for promoting this workshop.

References

[1] M. Cannataro, M. Romberg, J. Sundnes, R. W. dos Santos, Bioinformatics’ challenges to computer science, in: M. Bubak, G. D. van Albada, J. Dongarra, P. M. A. Sloot (Eds.), ICCS (3), Vol. 5103 of Lecture Notes in Computer Science, Springer, 2008, pp. 67–69.
[2] M. Cannataro, R. W. dos Santos, J. Sundnes, Bioinformatics’ challenges to computer science: Bioinformatics tools and biomedical modeling, in: G. Allen, J. Nabrzyski, E. Seidel, G. D. van Albada, J. Dongarra, P. M. A. Sloot (Eds.), ICCS (1), Vol. 5544 of Lecture Notes in Computer Science, Springer, 2009, pp. 807–809.

[3] M. Cannataro, R. W. dos Santos, J. Sundnes, Biomedical and bioinformatics challenges to computer science, Procedia Computer Science 1 (1) (2010) 931 – 933, iCCS 2010. doi:DOI: 10.1016/j.procs.2010.04.102.
URL http://www.sciencedirect.com/science/article/B986S-506HM1Y-3K/2/84767eaa9ec3557e0ee4926693ed895f