Journal of Bioinformatics and Computational Biology
Vol. 13, No. 3 (2015) 1502002 (4 pages)

DOI: 10.1142/S0219720015020023

Imperial College Press
www.icpress.co.uk

Introduction to the CSBio2014 special issue

Published 30 April 2015

The 5th International Conference on Computational Systems-Biology and Bioinformatics (CSBio2014) was hosted at Nanyang Technological University from 10th to 12th November 2014, and was jointly organised by Nanyang Technological University, National University of Singapore, Agency for Science, Technology and Research (A*STAR) and King Mongkut's University of Technology Thonburi.

The CSBio conference brought together researchers and practitioners to exchange ideas and stimulate research collaborations in response to rapid advances in the generation of high-throughput "omics" data (such as deep sequencing) and clinical data (e.g. biomedical images, and electronic health records). To extract novel biological knowledge from the heterogeneous data sources, and translate it into useful biological and clinical applications to benefit the society (e.g. better medicine and healthcare), novel computational techniques, tools, domain knowledge and insights are needed.

We kept to the tradition of carefully selecting a handful of high-quality papers to recommend for journal publications. This year, we teamed up with the *Journal of Bioinformatics and Computational Biology* (JBCB) to publish our selected papers. Particularly, we recommended 12 papers (out of 36 submissions) to JBCB, and 7 of these papers were finally accepted for publication. These papers can be categorized into two sections. The first section focuses on systems biology and medicine. The second section discusses sequence, structure and text-mining. We provide a brief description for each of them as follows.

Systems biology and medicine

Biological systems are composed of biomolecules such as genes, proteins, metabolites, and signaling components, which interact in complex networks. The paper "Investigating noise tolerance in an efficient engine for inferring biological regulatory networks" (Komori *et al.*, 2015) studies how to infer regulatory networks from experimental time series data. Particularly, the authors investigated and tested the noise tolerance of the proposed inferring engine when considering the error (noise) that is inherent in experimental data.

The paper "Computationally predicting protein-RNA interactions using only positive and unlabeled examples" (Cheng *et al.*, 2015) studies how to accurately

predict Protein-RNA interactions (PRIs), which are very important in a wide variety of cellular processes. Different from existing computational methods which artificially construct negative samples (which could have false negatives), this paper proposes a novel method PRIPU that employs biased-SVM for predicting protein-RNA interactions using only positive and unlabeled examples. Experimental results over three datasets show that the proposed method not only outperforms four existing methods, but also is able to predict unknown PRIs. Source code, datasets and related documents of PRIPU are available at http://admis.fudan.edu.cn/projects/pripu.html.

A major goal of personalized anti-cancer therapy is to increase drug effects while reducing side effects as much as possible. A novel therapeutic strategy called synthetic lethality (SL) provides a great opportunity to achieve this goal. A pair of genes is said to have SL interaction if mutations of both genes lead to cell death while mutation of either single gene does not. Hence, the SL partner of a gene mutated only in cancer cells could be a promising drug target, and the identification of SL pairs of genes is of great importance for the pharmaceutical industry. The paper "Predicting essential genes and synthetic lethality via influence propagation in signaling pathways of cancer cell fates" (Zhang *et al.*, 2015) presented a hybridized method to predict SL pairs of genes, which combined a data-driven model with knowledge of signaling pathways to simulate the influence of single-gene knock-down and double-gene knock-down to cell death. A validation using the literature shows that the predicted SL candidates agree well with wet-lab experiments and a few novel reliable SL candidates are also predicted by the proposed model.

Sequence, structure and text-mining

Sequence data repositories archive and disseminate fastq data in compressed format. The paper entitled "FQC: A novel approach for efficient compression, archival and dissemination of fastq datasets" (Dutta *et al.*, 2015) presents a novel technique FQC, a fastq compression method that provides significantly higher compression gains over current methods. In addition, the proposed technique also incorporates features necessary for universal adoption by data repositories or end-users. To make this research more easily adopted by academic data repositories, Linux, Windows and Mac implementations of FQC are freely available at http://metagenomics.atc.tcs.com/compression/FQC.

Metagenomics approach involves extraction, sequencing and characterization of the genomic content of the entire community of microbes present in a given environment. In contrast to genomic data, accurate assembly of metagenomic sequences is a challenging task due to the huge volume and the diverse taxonomic origin of metagenomic sequences. The paper "Grid-Assembly: An oligo-nucleotide composition based partitioning strategy to aid metagenomic sequence assembly" (Ghosh *et al.*, 2015) partitions metagenomic sequence datasets into clusters and assembles separately the sequences in individual clusters using any single-genome assembly method. Validation results indicate that the proposed approach helps in improving

the overall quality of assembly, in terms of the purity and volume of the assembled contigs.

It is well known that protein-ligand docking is an important step in drug discovery processes. The paper "PSOVina: The hybrid particle swarm optimization algorithm for protein-ligand docking" (Ng *et al.*, 2015) aims to accurately predict and efficiently optimize the positions and orientations of ligands in a binding pocket of a target protein. Particularly, this paper presents a new method PSOVina that combines particle-swarm optimization (PSO) algorithm with the efficient Broyden-Fletcher-GoldfarbShannon (BFGS) local-search method adopted in AutoDock Vina to tackle the conformational search problem in docking. Extensive experimental results show that, compared to the original Vina program, the proposed PSOVina achieves a remarkable speed-up of 40–44% in terms of average docking time without compromising prediction accuracies in the docking, making PSOVina a better choice as a docking tool in large-scale protein-ligand docking applications.

Molecular events normally have significant meanings since they describe important biological interactions or alternations such as the binding of a protein. As a crucial step of biological event extraction, event-trigger identification has attracted much attention where rule-based approach and machine-learning approach have been proposed. While machine-learning-based approaches have outperformed rule-based approaches, they still face a notable challenge, i.e. how to model semantic and syntactic information of different words and incorporate it into the prediction model. In the study "Embedding assisted prediction architecture for event trigger identification" (Nie *et al.*, 2015), a word-embedding-assisted neural-network prediction model has been proposed to conduct event trigger identification. The experimental results on a benchmark data set have shown its potential to offer researchers insights into semantic-aware solutions for event-trigger identification.

We thank all the authors of these papers on the exciting and important research topics. We also express our sincere gratitude to everyone involved in making the conference a success. Many thanks to advisory board members, the organizing committees, the keynote and plenary speakers, and particularly our conference General Chair Professor Chee-Keong Kwoh and General Co-Chair Professor Jonathan Chan. Our heartfelt thanks also go to PC members and the external reviewers for their hard work. Last but not least, we thank all the conference participants for attending and presenting at CSBio2014 in Singapore. It is our hope that the CSBio conference will provide a lasting platform for disseminating the latest research results and practice of the computational systems-biology and bioinformatics approaches in biology, medicine and healthcare.

References

1. A. Komori *et al.*, Investigating noise tolerance in an efficient engine for inferring biological regulatory networks, *Journal of Bioinformatics and Computational Biology* **13**(3):1541006, 2015.

2. Z. Cheng *et al.*, Computationally predicting protein-RNA interactions using only positive and unlabelled examples, *Journal of Bioinformatics and Computational Biology* **13**(3):1541005, 2015.
3. F. Zhang *et al.*, Predicting essential genes and synthetic lethality via influence propagation in signalling pathways of cancer cell fates, *Journal of Bioinformatics and Computational Biology* **13**(3):1541002, 2015.
4. A. Dutta *et al.*, FQC: A novel approach for efficient compression, archival and dissemination of FASTQ datasets, *Journal of Bioinformatics and Computational Biology* **13**(3):1541003, 2015.
5. T. S. Ghosh *et al.*, Grid-Assembly: An oligo-nucleotide composition based partitioning strategy to aid metagenomics sequence assembly, *Journal of Bioinformatics and Computational Biology* **13**(3):1541004, 2015.
6. M. C. K. Ng *et al.*, PSOVina: The hybrid particle swarm optimization algorithm for protein-ligand docking, *Journal of Bioinformatics and Computational Biology* **13**(3):1541007, 2015.
7. Y. Nie *et al.*, Embedding assisted prediction architecture for event trigger identification, *Journal of Bioinformatics and Computational Biology* **13**(3):1541001, 2015.

Xiao-Li Li and Jie Zheng
CSBio2014 PC Co-chairs