Chapter XXVIII
Multi–Level Data Integration and Data Mining in Systems Biology

Roberta Alfieri
CNR - Institute for Biomedical Technologies, Italy
and CILEA, Italy

Luciano Milanesi
CNR - Institute for Biomedical Technologies, Italy

ABSTRACT

This chapter aims to describe data integration and data mining techniques in the context of systems biology studies. It argues that the different methods available in the field of data integration can be very useful in making research in the field of systems biology easier. Moreover data mining is an important task to take into account in this context, therefore in this chapter, some aspects of data mining applied to systems biology specific case studies shall be discussed. The availability of a large number of specific resources, especially for the experimental researchers, is something difficult for users who tried to explore gene, protein, and pathway data for the first time. This chapter finally aims to highlight the complexity in the systems biology data and to provide an overview of the data integration and mining approaches in the context of systems biology using a specific example for the Cell Cycle database and the Cell Cycle models simulation.

INTRODUCTION

In the context of the application of biomedical science to systems biology, the availability of many different database and data resources, and a huge amount of heterogeneous data are continuously accumulating, became a crucial point in the last few years. In the field of the medical sciences, and, more

in particular in the systems biology context, it is largely recognized that successful data integration has become essential in order to improve the possibility to better explore the knowledge space in many different biological studies. Experimental researchers and computer scientists can discover through data integration new and interesting relationships that enable them to make better and faster decisions for example about disease targets and drug molecules. Moreover, the collection of related information has been shown to be an essential component in biomedical and systems biology research, particularly in the genomics, proteomics and pathways information area.

The necessity for data integration is widely approved in the bioinformatics and systems biology community since bioinformatics data are currently spread across the internet and throughout organizations in a wide variety of formats. Moreover the achievement of interesting results in most bioinformatics and systems biology-related activities, from functional characterization of genomic and proteomic data to the development of mathematical models of biological processes, requires an integrated view of all relevant data useful to accomplish those tasks. The challenges of data integration may be addressed using a wide variety of approaches. While each approach has advantages and limits, it can be difficult to evaluate which approach suits a particular need best without fully understanding the data integration landscape. The data integration methods aim to facilitate detailed and accurate investigation on specific gene, protein or pathway since high information content should be useful both for data mining and mathematical modelling of the biological process of interest. In this chapter the different data integration approaches and some practical example of data integration are illustrated in the specific field of the cell cycle process. The importance of the cell cycle in the shifting from a healthy to a pathological state in some specific experimental conditions which are illustrated in the context of the need to create an integrated system capable to collect the most important information related to cell cycle genes and proteins, which are drawn from the analysis of the cell cycle information available in literature and the existing pathway databases.

There is another important technique used for the knowledge discovery is the data mining approach. Data mining system has become widely used in the context of biomedical science and systems biology as it makes the prediction of the behaviours and the future trends for a biological system possible, allowing taking knowledge-driven decisions. In its general definition data mining can also be considered as the process of analyzing data from different perspectives and summarizing it into useful information, which can be used to increase the current knowledge about a specific biological process. Technically, data mining is the process of finding correlations or patterns among many fields in large relational databases. An example of data mining application in systems biology in the context of the mathematical modelling of a biological process is illustrated in this chapter.

Moreover the use of bioinformatic tools, data mining and data integration can help researchers to better studying the modelling complexity, by screening of the potential model components in order to find the emergent properties of a biological system, which is one of the main aims of systems biology studies. Finally the main advantage of using the data mining and the data integration approaches in the context of systems biology investigations, are presented.

BACKGROUND

During the past years, a very fast increase of the availability of quantitative data related to biological systems and processes occurred as consequence of the systematic application of automated high-

Related Content

Biomechanical Effects of Different Footwear on Steady State walking

Saad Jawaid Khan, Abu Zeeshan Bari, Soobia Saad Khanand Muhammad Tahir Khan (2016). *International Journal of Biomedical and Clinical Engineering (pp. 9-20).*

www.irma-international.org/article/biomechanical-effects-of-different-footwear-on-steady-state-walking/145163

GUI-CAD Tool for Segmentation and Classification of Abnormalities in Lung CT Image

V. Vijaya Kishoreand R.V.S. Satyanarayana (2019). *International Journal of Biomedical and Clinical Engineering (pp. 9-31).*

www.irma-international.org/article/gui-cad-tool-for-segmentation-and-classification-of-abnormalities-in-lung-ct-image/219304

Optimized Clustering Techniques with Special Focus to Biomedical Datasets

Anusuya S. Venkatesan (2018). *Biomedical Engineering: Concepts, Methodologies, Tools, and Applications (pp. 1149-1179).*

www.irma-international.org/chapter/optimized-clustering-techniques-with-special-focus-to-biomedical-datasets/186721

Application of Prewhitening Beamformer with Linear Constraints for Correlated EEG Signal Source Estimation

Teruyoshi Sasayama, Shoji Hamadaand Tetsuo Kobayashi (2013). *Biomedical Engineering and Cognitive Neuroscience for Healthcare: Interdisciplinary Applications (pp. 243-254).*

www.irma-international.org/chapter/application-prewhitening-beamformer-linear-constraints/69924

Study of Real-Time Cardiac Monitoring System: A Comprehensive Survey

Uma Arunand Natarajan Sriraam (2016). *International Journal of Biomedical and Clinical Engineering (pp. 53-63).*

www.irma-international.org/article/study-of-real-time-cardiac-monitoring-system/145167