

Chapter 34

PAGeneRN: Parallel Architecture for Gene Regulatory Network

Dina Elsayad

Ain Shams University, Egypt

Howida A. Shedeed

Ain Shams University, Egypt

A. Ali

Ain Shams University, Egypt

Mohamed F. Tolba

Ain Shams University, Egypt

ABSTRACT

The gene expression analysis is an important research area of Bioinformatics. The gene expression data analysis aims to understand the genes interacting phenomena, gene functionality and the genes mutations effect. The Gene regulatory network analysis is one of the gene expression data analysis tasks. Gene regulatory network aims to study the genes interactions topological organization. The regulatory network is critical for understanding the pathological phenotypes and the normal cell physiology. There are many researches that focus on gene regulatory network analysis but unfortunately some algorithms are affected by data size. Where, the algorithm runtime is proportional to the data size, therefore, some parallel algorithms are presented to enhance the algorithms runtime and efficiency. This work presents a background, mathematical models and comparisons about gene regulatory networks analysis different techniques. In addition, this work proposes Parallel Architecture for Gene Regulatory Network (PAGeneRN).

INTRODUCTION

Bioinformatics is the application of information technology in the field of molecular biology to manage process and analyze both genomic and molecular biological data. Its primary goal is to improve the understanding of biological processes (Nair, 2007). Bioinformatics is a very rich filed of research, in fact it includes so many research areas with diverse applications and according to Srikanth Aluru (2006), the major research areas in Bioinformatics are: Genome annotation (Kelley, MacCallum, & Sternberg, 2000), Sequence analysis (Durbin 1998, Sharif et al. 2015, Sharif et al. 2016, Tharwat et al. 2015),

DOI: 10.4018/978-1-5225-2229-4.ch034

Computational evolutionary biology (Cosmides & Tooby, 1995), Gene expression analysis (Velculescu, Zhang, Vogelstein, & Kinzler, 1995), Analysis of regulation (Janssen, Jones, Jones, & Woods, 1988) and Protein expression analysis (Ghaemmaghami, et al., 2003). In addition, it includes analysis of mutations in cancer (Berrozpe, Schaeffer, Peinado, Real, & Perucho, 1994), Predictions of protein structure (Shortle, 2000), Comparative Genomics (Rubin, et al., 2000), Modeling biological systems (Haefner, 2005), High throughput image analysis (Dowsey, 2005), Protein-protein docking (Dominguez, Boelens, & Bonvin, 2003) and Microarrays (Churchill, 2002).

The microarray is one of the active research fields in bioinformatics. It's a multiplex lab-on-a-chip that assays large amounts of biological material using high throughput screening methods (Culf, Cuperlovic-Culf, & Ouellette, 2006). Moreover, the microarrays experiment (Li & Li, 2008) starts with the sample preparation after that the hybridization, diction step and finally the data acquisition and analysis step. The microarray chips are scanned using special devices called Microarray Scanner. The microarray type depends on the biological material that assays. There are a number of types of microarrays including DNA microarrays (Gasch, et al., 2000), MMChips, for surveillance of microRNA populations (Meenakshisundaram, et al., 2009), Protein microarrays (Stoevesandt, Taussig, & He, 2009), Tissue microarrays (Kononen, et al., 1998), Cellular microarrays (Ma & Horiuchi, 2006), Chemical compound microarrays (Ma & Horiuchi, 2006), Antibody microarrays (Rivas, et al., 2008) and Carbohydrate arrays (glycoarrays) (Culf, Cuperlovic-Culf, & Ouellette, 2006).

The DNA Microarray technology is a high throughput experimental technique that can measure expression levels of hundreds of thousands of genes simultaneously. As an illustration, expression level of the gene is estimated by measuring the amount of mRNA for that gene. A gene is active if it is being transcribed; the more mRNA usually indicates more gene activity (Schena, Shalon, Davis, & Brown, 1995). The DNA microarray has become a useful technique in gene expression analysis for the development of new diagnostic tools and for the identification of disease genes and therapeutic targets for disease such as cancer. In fact, it's the basis for a very efficient screening and diagnosis of disease in an early stage of development. Microarrays could also be expanded into a starting point for developing new treatments for various types of disease.

The analysis of the data resulting from microarrays remains a big challenge for the huge volume of data it produces. In addition, the data analysis process involves various computational tasks (Yang, et al., 2008) and one of the microarrays data analysis tasks is network analysis (Haman & Valenta, 2013). The gene regulatory network (Aluru, 2006) goal is to study the topological organization of the genes interactions. The gene regulatory network analysis techniques can be classified into four main categories: component analysis techniques (Raychaudhuri, Stuart, & Altman, 2000) (Watkins, 2004) (Holter, et al., 2000) (Hyvarinen, Karhunen, & Oja, 2001) (Aapo, 1999) (Liebermeister, 2002) (Liao, et al., 2003) (Chang, Ding, Hung, & Fung, 2008), mutual information based techniques (Romualdi & Chiara, 2011) (Shi, Schmidt, Liu, & Muller-Wittig, 2011), reverse engineering techniques (Jostins & Jaeger, 2010)[44] and the regression techniques (Gregoretti, Belcastro, Di Bernardo, & Oliva, 2010). In fact, almost all of these techniques are computational intensive and time-consuming. Therefore, parallel algorithms are needed to enhance the techniques performance. The parallelism techniques are categorized as shared memory (Karp, 1988) (Hager & Wellein, 2010) and distributed memory (Hager & Wellein, 2010). Moreover, these two models can be mixed in the hybrid model to take the full advantage of the computing power (Hager & Wellein, 2010). Another form of the distributed computing is cloud computing (Erl, Puttini, & Mahmood, 2013). In addition, the graphic processing unit (GPU) can be used to enhance performance.

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/pagenern/180972

Related Content

Characterization and Modelization of Surface Net Radiation through Neural Networks

Antonio Geraldo Ferreira, Emilio Soria, Antonio J. Serrano López and Ernesto Lopez-Baeza (2010). *Soft Computing Methods for Practical Environment Solutions: Techniques and Studies* (pp. 260-277).

www.irma-international.org/chapter/characterization-modelization-surface-net-radiation/43156

Simulating Timing Behaviors for Cyber-Physical Systems Using Modelica

Hao Zhou, Mengyao Zhao, Linbo Wu and Xiaohong Chen (2019). *International Journal of Software Science and Computational Intelligence* (pp. 44-67).

www.irma-international.org/article/simulating-timing-behaviors-for-cyber-physical-systems-using-modelica/236151

A Fig-Based Method for Prediction Alumina Concentration

Jun Yi, Jun Peng and Taifu Li (2012). *International Journal of Software Science and Computational Intelligence* (pp. 41-50).

www.irma-international.org/article/a-fig-based-method-for-prediction-alumina-concentration/88926

Improved SfM-Based Indoor Localization with Occlusion Removal

Yushi Li, George Baci, Yu Han and Chenhui Li (2018). *International Journal of Software Science and Computational Intelligence* (pp. 24-40).

www.irma-international.org/article/improved-sfm-based-indoor-localization-with-occlusion-removal/207743

Telemetry Data Mining Techniques, Applications, and Challenges

Sara Ahmed, Tarek Gaber and Aboul Ella Hassanien (2017). *Handbook of Research on Machine Learning Innovations and Trends* (pp. 915-924).

www.irma-international.org/chapter/telemetry-data-mining-techniques-applications-and-challenges/180978