

Inferring Gene Regulatory Networks: Challenges and Opportunities

Jason T. L. Wang*

Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102, USA

*Corresponding author: Jason T. L. Wang, Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102, USA, Tel: 973-596-3396; Fax: 973-596-5777; E-mail: wangj@njit.edu

Rec date: Dec 22, 2014; Acc date: Dec 26, 2014; Pub date: Jan 2, 2015

Copyright: © 2015 Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Editorial

With the development of high-throughput, next-generation sequencing and other advanced technologies, a large number of gene expression profiles have been produced. Many of these profiles are available from public databases [1-3]. A challenging research problem that has drawn a lot of attention in the past is to infer gene regulatory networks from the expression data. A gene regulatory network is represented by a directed graph, in which nodes represent transcription factors or mRNA with edges showing transcriptional regulatory relationships between two nodes.

Maetschke et al. [4] categorized existing network inference methods into three groups: unsupervised, supervised and semi-supervised. While supervised algorithms are capable of achieving the highest accuracy among all the network inference methods, these algorithms require a large number of positive and negative training examples. An example here refers to an edge between two nodes in a network. A positive example refers to a known interaction between two genes while a negative example refers to an interaction that is known not to exist between two genes. Negative examples are difficult to obtain in many organisms. Instead, some researchers use unknown interactions between genes for negative examples. Unsupervised algorithms infer networks based solely on gene expression profiles and do not need any training examples. The accuracy of these algorithms is usually low. However, these algorithms are useful for organisms where training data are not available. Semi-supervised algorithms often exploit positive-unlabeled (PU) learning techniques by taking a small sample of positive examples and a large number of unlabeled examples to train a classification model and use the trained model to predict a network.

Marbach et al. [5, 6] developed an in silico benchmark suite within the DREAM (Dialogue on Reverse Engineering Assessment and Methods) project [7, 8], and assessed the performance of 29 network inference methods. They concluded that reliable network inference from gene expression data remains an unsolved problem. Madhamshettiwar et al. [9] evaluated nine state-of-the-art gene regulatory network inference methods using 38 simulated datasets. These authors observed that the performance of the evaluated methods depends on many factors such as features of the data, network size and topology, as well as parameter settings. Indeed, the parameter settings often affect the accuracy of a network inference method, and identifying the optimal parameter values is a very challenging task.

Recently the network inference problem has been approached from many different perspectives, including:

- using hybrid and pipeline techniques to enhance the performance of unsupervised methods [10];

- employing advanced data mining and machine learning algorithms to improve the accuracy of supervised or semi-supervised methods [11];
- using new techniques to tackle constrained or condition-specific network inference problems [12, 13]; and
- adopting GPU and cloud technologies to speed up the execution of network inference algorithms [14].

These different directions are presenting new opportunities for researchers working in this field.

References

1. Petryszak R, Burdett T, Fiorelli B, Fonseca NA, Gonzalez-Porta M, et al. (2014) Expression Atlas update--a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. *Nucleic Acids Res* 42(Database issue): p. D926-932.
2. Hubble J, Demeter J, Jin H, Mao M, Nitzberg M, et al. (2009) Implementation of GenePattern within the Stanford Microarray Database. *Nucleic Acids Res* 37(Database issue): p. D898-901.
3. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. (2013) NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* 41(Database issue): p. D991-995.
4. Maetschke SR, Madhamshettiwar PB, Davis MJ, Ragan MA (2014) Supervised, semi-supervised and unsupervised inference of gene regulatory networks. *Brief Bioinform* 15: 195-211.
5. Marbach D, Prill RJ, Schaffter T, Mattiussi C, Floreano D, et al. (2010) Revealing strengths and weaknesses of methods for gene network inference. *Proc Natl Acad Sci U S A* 107: 6286-6291.
6. Schaffter T, Marbach D, Floreano D (2011) GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods. *Bioinformatics* 27: 2263-2270.
7. Prill RJ, Marbach D, Saez-Rodriguez J, Sorger PK, Alexopoulos LG, et al. (2010) Towards a rigorous assessment of systems biology models: the DREAM3 challenges. *PLoS One* 5: e9202.
8. Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, et al. (2012) Wisdom of crowds for robust gene network inference. *Nat Methods* 9: 796-804.
9. Madhamshettiwar PB, Maetschke SR, Davis MJ, Reverter A, Ragan MA (2012) Gene regulatory network inference: evaluation and application to ovarian cancer allows the prioritization of drug targets. *Genome Med* 4: 41.
10. Lingeman JM, Shasha D (2012) *Network Inference in Molecular Biology: A Hands-on Framework*, Springer, New York.
11. Gillani Z, Akash M, Rahaman M, Chen M (2014) CompareSVM: supervised, Support Vector Machine (SVM) inference of gene regulatory networks. *BMC Bioinformatics* 15: 395.
12. Cai X, Bazerque JA, Giannakis GB (2013) Inference of gene regulatory networks with sparse structural equation models exploiting genetic perturbations. *PLoS Comput Biol* 9: e1003068.
13. Ellwanger DC, Leonhardt JF, Mewes HW3 (2014) Large-scale modeling of condition-specific gene regulatory networks by information integration and inference. *Nucleic Acids Res* 42: 0.

14. Lee WP, Hsiao YT, Hwang WC (2014) Designing a parallel evolutionary algorithm for inferring gene networks on the cloud computing environment. BMC Syst Biol 8: 5.