Simulate Gene Expression and Infer Gene Regulatory Networks

Subjects: Computer Science, Artificial Intelligence Contributor: Francesco Zito , Vincenzo Cutello , Mario Pavone

The ability to simulate gene expression and infer gene regulatory networks has vast potential applications in various fields, including medicine, agriculture, and environmental science. Machine learning approaches to simulate gene expression and infer gene regulatory networks have gained significant attention as a promising area of research.

reverse engineering

gene regulatory network

machine learning

1. Introduction

Understanding the intrinsic relationship between genes with the aim of treating known diseases is currently one of the great challenges in genetics ^[1]. Although this topic may seem to be only a biological problem, it actually involves many areas of computer science. Due to the complexity of this problem, traditional mathematical methods such as Ordinary Differential Equations (ODEs), which rely on estimates of gene expression levels over time through a continuous model, may be inaccurate for a larger number of genes and require high-quality data to create an acceptable model ^[2]. Machine-learning-based techniques have emerged as a promising approach for gene regulatory network inference, outperforming other methods based on mutual information ^[3]. These techniques can be broadly classified into two categories. The first category involves using observations to create a model that approximates the real system, which is then used to construct a complex network that identifies the regulatory genes for other genes, known as the gene regulatory network ^[4]. The second category involves the direct creation of a gene regulatory network through observations, without the need to estimate a model representing the dynamics of gene expression ^[3].

Related Work

Before the spread of machine learning in the field of genetics, Boolean networks were generally used to describe gene regulatory networks. All biological components can be described by binary states and their interactions by Boolean functions ^[6]. Boolean networks are relatively simple to implement, but their implementation requires noise-free, discrete data, which can be difficult to obtain when working with real-world data ^[7].

In recent years, several methods to extract a gene regulatory network have been presented. In ^[B], the authors divided the methods for inferring a gene regulatory network from gene expression data into three main groups: (*i*) model-based methods; (*ii*) information-theory-based methods; and (*iii*) machine learning methods. Some

experimental tests have shown that machine learning methods can obtain a high accuracy in predicting gene interactions ^[2]. One approach to inferring a gene regulatory network from a gene expression dataset is to use differential equations. This requires a mathematical model of the changes in gene expression over time using ordinary differential equations, which can provide insight into the underlying dynamics of the system. By analyzing the behavior of these equations, one can gain a better understanding of how genes interact and regulate each other within the network ^[10]. The difficulty in such an approach is, clearly, to build a differential equation model from data. To this end, several methods have been proposed in the literature. An example can be found in ^[11], where a metaheuristic was used to find the parameters of an *S-system model* describing the dynamics of gene expression. Another example can be found in ^[12], where a complex-valued ordinary differential equation model was created using genetic programming. In addition, it is possible to directly predict the interaction between genes using a gene expression dataset. One method used in this field is GENIE3 presented in ^[13] and its improvement called DynamicGENIE3 presented in ^[14]. An improvement on the previously cited method in this category can be found in ^[15], where different inference methods are combined to increase the accuracy of the resulting gene regulatory network.

Rather than using a specific strategy to predict each arc of a gene regulatory network, an alternative approach involves the construction of a comprehensive network that assumes all possible interactions between genes, represented as a strongly connected graph, and subsequently applying a pruning strategy to eliminate non-corresponding arcs. One example of such a method, which employs an information-theoretic algorithm, is described in ^[16].

2. Simulate Gene Expression and Infer Gene Regulatory Networks

The process by which the instructions in our *DNA* are transformed into a functioning product, such as a protein, is known as *gene expression* ^[17]. Gene expression allows a cell to respond to changes in its environment. The *regulation of gene expression* (or just *gene regulation*) is a very complex process that takes into account several biological factors to respond, for example, to environmental stimuli or to adapt to new food sources ^{[18][19]}. Gene regulation involves a variety of mechanisms used by cells to increase or decrease the production of certain gene products. Thus, it functions like an on/off switch that regulates the amount of proteins produced.

Considering the huge amount of gene products that are present in a multicellular organism, the regulatory mechanisms are represented in a directed graph, called the *regulatory network*, to help better understand the regulatory mechanisms. A regulatory network reveals the interactions between genes, proteins, mRNAs, and cellular processes and provides important information about the development of diseases ^[20]. Knowledge of a regulatory network for an entire organism or for a small group of genes is crucial for a full understanding of the life process of an organism and how gene products interact with each other ^[21]. Once this is clear, it is possible to send external chemical signals to inhibit a gene that could be dangerous to the life of an organism, such as the development of a cancer cell or a genetic disease ^[22].

2.1. Gene Regulatory Network

A gene regulatory network is a directed graph where the nodes represent genes, and the directed arcs model the interactions between the genes ^[23]. Specifically, a *Gene Regulatory Network* (GRN) represents the regulatory process of gene expression in an organism. An arc between two nodes, i.e., genes, mainly provides information about the regulatory process. In the context of inferring gene regulatory networks, the presence of a direct arc from gene *Gi* to gene *Gj* indicates that *Gi* is a regulatory gene, also known as a regulator ^[24]. This implies that any alteration in the expression of *Gi* will have a consequential impact on the expression of *Gj*, according to the principle of cause and effect. In other words, the regulatory gene *Gi* is capable of influencing the expression of its target gene *Gj*, thereby establishing a cause-and-effect relationship between the two genes.

A gene regulatory network can, therefore, combine more-detailed regulatory information. In fact, a regulatory gene controls the expression of its associated genes in a positive or negative way. When the expression level of the regulator reaches a threshold, another gene can be activated or inhibited based on that level ^[25]. This results in a change in the expression level of the regulated gene: if the gene expression decreases, the gene is inhibited; otherwise, it is activated. **Figure 1** shows an example of a gene regulatory network. As can be seen, there are two types of arcs in a gene regulatory network: activation arcs and inhibition arcs.





2.2. Inferring a Gene Regulatory Network

The process of inferring a gene regulatory network for a cellular organism can be divided into four distinct phases, which researchers label: observation; modeling; inference; and validation. The whole process is shown in **Figure 2**.



Figure 2. Process to infer a gene regulatory network.

- **Observation:** The first step is to observe how the gene expression of a group of genes responds to external perturbations in a real organism. This can be performed using various strategies, such as microarray technology ^[26]. The level of gene expression for each gene is recorded over time to create a time-series dataset containing gene expression for the genes under observation. Typically, such a dataset is represented as a matrix $D \in \mathbb{R}^{M \times N}$, where *N* is the number of genes and *M* is the number of observations for each gene over time.
- **Modeling:** The gene expression time-series dataset is used to train a model that can be based on differential equations ^[27] or design an artificial environmental setting.
- **Inference:** The model created in the previous phase is used to make predictions about the relationships between genes in order to discover regulatory genes. This information can, therefore, be used to draw a complex network, i.e., a gene regulatory network, showing these relationships.
- Validation: Finally, to validate the accuracy of a predicted gene regulatory network, it is essential to compare it with the target network. However, this comparison can only be performed on an artificial dataset where the gene regulatory network is known beforehand. In a real organism, researchers do not have access to a gene regulatory network, and therefore, the validation of the predicted gene regulatory network must be performed empirically and in the field.

References

 Gout, J.F.; Kahn, D.; Duret, L.; Paramecium Post-Genomics Consortium. The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. PLoS Genet. 2010, 6, e1000944.

- 2. Karlebach, G.; Shamir, R. Modeling and analysis of gene regulatory networks. Nat. Rev. Mol. Cell Biol. 2008, 9, 770–780.
- 3. Shu, H.; Zhou, J.; Lian, Q.; Li, H.; Zhao, D.; Zeng, J.; Ma, J. Modeling gene regulatory networks using neural network architectures. Nat. Comput. Sci. 2021, 1, 491–501.
- 4. Aubin-Frankowski, P.C.; Vert, J.P. Gene regulation inference from single-cell RNA-seq data with linear differential equations and velocity inference. Bioinformatics 2020, 36, 4774–4780.
- Pratapa, A.; Jalihal, A.P.; Law, J.N.; Bharadwaj, A.; Murali, T.M. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. Nat. Methods 2020, 17, 147– 154.
- 6. Schwab, J.D.; Kühlwein, S.D.; Ikonomi, N.; Kühl, M.; Kestler, H.A. Concepts in Boolean network modeling: What do they all mean? Comput. Struct. Biotechnol. J. 2020, 18, 571–582.
- 7. Delgado, F.M.; Gómez-Vela, F. Computational methods for Gene Regulatory Networks reconstruction and analysis: A review. Artif. Intell. Med. 2019, 95, 133–145.
- 8. Zhao, M.; He, W.; Tang, J.; Zou, Q.; Guo, F. A comprehensive overview and critical evaluation of gene regulatory network inference technologies. Briefings Bioinform. 2021, 22, bbab009.
- 9. Pirooznia, M.; Yang, J.Y.; Yang, M.Q.; Deng, Y. A comparative study of different machine learning methods on microarray gene expression data. BMC Genom. 2008, 9 (Suppl. S1), S13.
- Cao, J.; Qi, X.; Zhao, H. Modeling Gene Regulation Networks Using Ordinary Differential Equations. In Next Generation Microarray Bioinformatics: Methods and Protocols; Wang, J., Tan, A.C., Tian, T., Eds.; Humana Press: Totowa, NJ, USA, 2012; pp. 185–197.
- Agostini, D.; Costanza, J.; Cutello, V.; Zammataro, L.; Krasnogor, N.; Pavone, M.; Nicosia, G. Effective calibration of artificial gene regulatory networks. In Proceedings of the 2011 11th European Conference on Artificial Life (ECAL), Paris, France, 8–12 August 2011; p. 11.
- Yang, B.; Bao, W.; Zhang, W.; Wang, H.; Song, C.; Chen, Y.; Jiang, X. Reverse engineering gene regulatory network based on complex-valued ordinary differential equation model. BMC Bioinform. 2021, 22, 448.
- 13. Huynh-Thu, V.A.; Irrthum, A.; Wehenkel, L.; Geurts, P. Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. PLoS ONE 2010, 5, e12776.
- 14. Huynh-Thu, V.A.; Geurts, P. dynGENIE3: Dynamical GENIE3 for the inference of gene networks from time-series expression data. Sci. Rep. 2018, 8, 3384.
- 15. Åkesson, J.; Lubovac-Pilav, Z.; Magnusson, R.; Gustafsson, M. ComHub: Community predictions of hubs in gene regulatory networks. BMC Bioinform. 2021, 22, 58.

- 16. Hartemink, A.J. Reverse engineering gene regulatory networks. Nat. Biotechnol. 2005, 23, 554– 555.
- Emmert-Streib, F.; Dehmer, M.; Haibe-Kains, B. Gene regulatory networks and their applications: Understanding biological and medical problems in terms of networks. Front. Cell Dev. Biol. 2014, 2, 38.
- Emerson, J.J.; Li, W.H. The genetic basis of evolutionary change in gene expression levels. Philos. Trans. R. Soc. B Biol. Sci. 2010, 365, 2581–2590.
- 19. Davidson, E.; Levin, M. Gene regulatory networks. Proc. Natl. Acad. Sci. USA 2005, 102, 4935.
- 20. Glubb, D.M.; Innocenti, F. Mechanisms of genetic regulation in gene expression: Examples from drug metabolizing enzymes and transporters. WIREs Syst. Biol. Med. 2011, 3, 299–313.
- Huynh-Thu, V.A.; Sanguinetti, G. Gene Regulatory Network Inference: An Introductory Survey. In Gene Regulatory Networks: Methods and Protocols; Sanguinetti, G., Huynh-Thu, V.A., Eds.; Springer: New York, NY, USA, 2019; pp. 1–23.
- Zhang, Z.; Lei, A.; Xu, L.; Chen, L.; Chen, Y.; Zhang, X.; Gao, Y.; Yang, X.; Zhang, M.; Cao, Y. Similarity in gene-regulatory networks suggests that cancer cells share characteristics of embryonic neural cells. J. Biol. Chem. 2017, 292, 12842–12859.
- 23. Vijesh, N.; Chakrabarti, S.K.; Sreekumar, J. Modeling of gene regulatory networks: A review. J. Biomed. Sci. Eng. 2013, 6, 9.
- 24. Hecker, M.; Lambeck, S.; Toepfer, S.; van Someren, E.; Guthke, R. Gene regulatory network inference: Data integration in dynamic models—A review. Biosystems 2009, 96, 86–103.
- 25. Wang, Y.R.; Huang, H. Review on statistical methods for gene network reconstruction using expression data. J. Theor. Biol. 2014, 362, 53–61.
- 26. Müller, U.R.; Nicolau, D.V. Microarray Technology and Its Applications; Springer: Berlin/Heidelberg, Germany, 2005.
- 27. Gebert, J.; Radde, N.; Weber, G.W. Modeling gene regulatory networks with piecewise linear differential equations. Eur. J. Oper. Res. 2007, 181, 1148–1165.

Retrieved from https://encyclopedia.pub/entry/history/show/109952