Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/13/20230738

Computational Methods for Inferring and Reconstructing Gene Regulatory Networks

Xianglu Zhou[1,4], Haiyang Jia[1,2,3,5,*][0000-0003-0951-6671]

[1]College of Computer Science and Technology, Jilin University, Changchun 130012, China
[2]College of Software, Jilin University, Changchun 130012, China
[3]Key Laboratory of Symbolic Computation and Knowledge, Engineering of Ministry of Education, Jilin University, Changchun 130012, China

[4]astin001031@icloud.com
[5]jiahy@jlu.edu.cn

Abstract. Gene regulatory network (GRN) describes dynamic and complex gene interactions that determine the functions of cells. Ergo, understanding of GRNs has played a great role in cancer treatment, drug design, and gene research. However, the GRN study suffers from a lack of experiment measurement data and the complexity of the model. Computational approaches have evolved in recent years to promote the study of GRN. In this review, we summarized popular Computational approaches for GRN reconstruction. We started with traditional bioinformatic methods, such as Bayesian networks and mutual information methods. Later, we introduced how today's hot technology in the computer field - machine learning benefited GRN research. Tree-based approaches and other machine-learning methods are elaborated on in this section. We discussed not only the advantages and progression brought by various methods but also the drawbacks and limitations, such as the accuracy and robustness of GRN reconstruction. We expect to inspire readers for improved GRN study approaches via our introduction to this field.

Keywords: Gene Regulatory Network, Bioinformatics, Machine Learning.

1. Introduction

A gene regulatory network (GRN) is a group of molecular regulators that work together in cells to control the levels of mRNA and protein mRNA production, which in turn controls the specific function of the cells. Therefore, GRNs play a key role in the formation of body structures, morphogenesis, and evolutionary developmental biology. The regulator can be DNA, RNA, protein, or any combination of these three to form a complex and the interaction can be of different organizations. A major group of regulator proteins is the transcription factors, which play a main role in regulatory networks or cascades by regulating the expression of different groups of genes with diverse functions. By binding on the promoter region at the start of genes, transcription factors can either initiate the expression of target genes or inhibit gene expression in rare cases [1]. Transcriptional factor-mediated-GRN, together with other regulators, chooses which genetic functions will be expressed in each geographic region of the organism's cells at every developmental stage [2].

DOI: 10.54254/2755-2721/13/20230738

Many life science fields, with a particular focus on basic and applied biological research, medical science, and drug discovery, are greatly impacted by understanding the regulations of gene expression under various conditions. This is because many dysregulated gene expression programs are at the root of a wide variety of diseases [3]. Understanding the molecular pathways governing healthy cells and disease states is essential for finding new possible treatment targets. Precise GRN inference has shown to be a trustworthy technique in the investigation of critical cellular systems activities such as development, differentiation, metabolism, adaptability, and signaling. Many applications of GRNs provide opportunities for research fields such as plant scienceand human health to make innovations [4-8]. The reconstruction and analysis of GRNs are thought to be fundamental for a better understanding of the mechanisms underlying the expression of genes and the effect of their perturbation in the context of specific biological mechanisms or pathways. Additionally, the diagnostics and treatment of various types of cancers and other human diseases were greatly aided by the pertinent studies of GRNs (especially transcription factor-guided GRNs) [9-11].

However, reconstruction and inference of GRN have remained a great challenge. Due to the dynamic and complex nature of GRNs and the mismatch between observed measurements and targeted networks, the accuracy, such as indirect interactions and biological closeness [12,13], of the reconstructed GRN is always a big concern in this field. Also, the lack of data is a consistent problem for human disease GRN constructions [14].

During the past decades, the rapid development of high-throughput technologies provides great opportunities for GRN study. Next Generation Sequencing (NGS) [15], which provides a leading look into RNA and DNA samples, has significantly improved in robustness, quality, and low noise. The analysis of the dynamic cell transcriptome and the hunt for tracking changes in gene expression in cells or organs under various situations are made possible by the widespread use of microarray and RNA-Seq (RNA sequencing) [16,17], which has produced a wealth of diverse data on gene co-expression. Also, experimental approaches such as Chromatin Immunoprecipitation sequencing (ChIP-seq) were developed to capture protein-DNA interactions directly [18], which helps to recover one-to-multiple relations between transcription factors and their target genes efficiently. Together, these approaches provide rich information for us to understand gene regulation and GRN.

Along with the progress of experimental approaches, various computational tools were developed to process and interpret the rich but complex information derived from experiments. Boolean networks, ordinary differential equations (ODE), Bayesian networks, and mutual information (MI)-based approaches are widely known and commonly used in reconstructing GRNs with sequencing data. They vary in resolutions and show different advantages for different purposes, while all following a paradigm of reverse engineering [19].

Recently, machine learning and deep learning-based methods are also proposed for GRN studies. Machine learning (ML) 'learns' general rules to perform assigned tasks (such as GRN reconstruction) without being explicitly programmed to do so by leveraging sample data, known as training data [20,21]. Deep learning is part of a broader family of machine learning methods based on artificial neural networks whose 'learning' process is more complex and often black-boxed [9]. With reduced reliance on past information and greater efficiency when compared to conventional methods, machine learning, and deep learning-based methods can better accommodate complicated relationships between transcription factors and the genes they regulate [22]. Though they also face problems such as the huge need for data.

In this paper, we summarized the principles, advantages and disadvantages, tools, and applications of current computational approaches widely used for transcription factor-based GRN reconstruction. We first introduced traditional approaches widely used for GRN study, while listing their differences and preferred scenarios. Next, we discussed the application of machine learning algorithms in the GRN study. We highlighted decision tree-based approaches, a machine learning approach prevalent in the GRN study. Also, we compared these machine learning methods with traditional methods. Pointing out the limitations and opportunities of current approaches, we expect improved approaches can further be developed to promote GRN reconstruction.

DOI: 10.54254/2755-2721/13/20230738

2. Traditional Approaches

In this section, we introduced the four most commonly used GRN reconstruction approaches with examples of implementations.

In Boolean networks, gene expression levels are discretized into Boolean binary values, with 0 denoting silent or almost silenced genes and 1 denoting activated genes in boolean networks, which represent genes as variables. On/off expression values are used to represent genes, and logic operators are used to connecting the Boolean variables in a Boolean function [23]. As gene expression rarely involves either complete activation or complete silence, some crucial features might be missed. Barman et al. developed a Boolean network model named MIBNI to identify initial regulatory genes in E. coli and fission yeast cell cycle networks with time-series gene expression data [24]. They leveraged the logical rule commonly used in the Boolean network but also introduced mutual information-based feature selections to enable large-scale network inference.

In contrast to Boolean networks, ordinary differential equations treat changes in gene expression as a function of certain other genes expressions, treat changes in gene expression as a function of environmental considerations, and use continuous variables rather than discrete variables to produce more accurate and precise models that are more comparable to the real behavior of the biological process [25].

Inferelator is an ODE-based approach that interprets transcriptional expression profiles (TEPs) under several experimental interferences [26,27]. TEP can be used to determine whether two genes are co-expressed when the TEP of the two genes has obvious dependence [28]. Inferelator learns a system of simplified ODEs by introducing prior knowledge from artificial gene annotation and L1-constrained network architecture. This ODE system is set up to approximate decoupling, and then implemented using a one-step finite difference for each ODE solution Expanding on this, the system contains functions as environmental and transcription factors and the transcriptional change rate of genes, which significantly improved the ability to model longer time scales in performance.

Bayesian networks are probabilistic models that incorporate Bayesian models, probability theory, and graph theory. A directed acyclic plot, which also represents the local joint probability distribution of all interactions or the relationship between all transcripts recorded in the same sample, is used by Bayesian networks to represent the conditional dependencies of a set of random variables [29]. When extrapolating dynamic GRNs, Bayesian approaches are frequently regarded as the best approach. These methods have applications in a wide range of disciplines, including medicine and evolutionary development. Dynamic Bayesian networks are an extension of Bayesian networks that use probabilistic graphical models to infer uncertainty in interactions between genes [30]. In contrast to Bayesian networks, dynamic Bayesian networks are able to simulate circulatory interactions between genes and are an important aspect of biological network modeling [31]. Vinh et al. reconstructed gene regulatory networks using time-course data sets by combining a dynamic Bayesian network approach with deterministic global optimization techniques [32]. The approach named GlobalMIT+ learned high-order time-delayed genetic interactions accurately when evaluating both synthetic and real datasets.

Compared with Boolean networks, Bayesian networks can make better dynamic GRNs due to the combination of different types of data and prior knowledge and can achieve more accurate systems like ODE methods. And while Boolean networks use discrete variables and ODE methods use continuous variables, Bayesian networks can use discrete or continuous expression levels during the learning process.

A measurement of the interdependence between two variables is the mutual information (MI) of two random variables. More specifically [33], it measures how much "knowledge" can be learned about another random variable by seeing one. All gene pair associations can be inferred using MI in the context of GRN reconstruction. The ability of MI to infer nonlinear relationships between TEPS is its most significant advantage [34]. A conditional mutual information estimator based on adaptive division is proposed, allowing us to perform conditional estimation of discrete and continuous random variables simultaneously. This estimator uses mutual information and conditional mutual information to evaluate the interaction between genes [35]. Mutual information methods can find large GRNs from

DOI: 10.54254/2755-2721/13/20230738

low-expression data while the computational complexity is low. These techniques, however, are static and do not account for numerous genes that are engaged in a single regulation.

ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks) is a famous algorithm, which is based on mutual information. And it is designed to reverse transcription networks from transcriptome gene expression data [36]. By quantifying MI between expression profiles, this technique locates TF-TG gene pairs with comparable transcriptional fluctuations [12]. In the first step, ARACNe calculates all pairwise TF-TG mutual information values based on a gene expression dataset and a list of TFs. Then according to the set global salience threshold decide whether to keep the edge or not. In the second step, ARACNE uses a well-known information-theory property, which is data processing inequalities (DPI), eliminating most indirect candidate interactions. After obtaining the side information composed of three genes, the algorithm detects all three mutual information values, finds the gene triple group that is greater than the set threshold, and finally deletes the side with the smallest value.

Despite the fact that associative network-based techniques, such as ARACNE, are effective when interactions in gene regulatory networks happen between gene pairs, they are unable to fully uncover the relationships of target genes that are co-regulated by multiple genes. One such interaction is the XOR interaction rule, which can only be discovered by taking advantage of the conditional dependency between relevant variables. To find XOR and other nonlinear interactions between two genes, conditional mutual information (CMI) can be employed. MI and CMI-based methods, which are introduced in some research [37,38], reduce the false positive rate of detection interactions. Furthermore, the CMI-based inference method outperforms ARACNE when the underlying regulatory network incorporates co-regulatory or cross-regulatory genes [35].

In fact, the combination of the Bayesian network and MI has been widely used in GRNs. By performing CMI to integrate local BNS to form an exploratory network, or GRN, redundant rules in GRNs are reduced, thereby alleviating the false positive problem. Based on a test network, CMI, and local BN can be executed repeatedly to produce the final network or GRN. Over the course of iteration, fake or redundant rules are gradually removed [39]. Since different approaches provide information from different perspectives, the combination of different methods may improve the accuracy of GRN reconstruction.

3. Decision Tree-based Methods and Other Methods

Compared with traditional methods, in the field of GRN research, machine learning can be used to do more work. While it can be used as an inference algorithm, it can also be used for the reconstruction of GRN and the clustering of gene expression data. Moreover, deep learning is better at inferring complex, interdependent relationships between biological entities and processes due to its compositional ability to model nonlinear dependencies and is better at dealing with data noise and multi-scale problems. We will therefore concentrate on decision tree-based methods before moving on to other methods.

3.1. Decision Tree-based Methods

A group of if-then rules is utilized by tree-based models to produce predictions from one or more decision trees. The capability of tree-based approaches to identify multivariate interaction effects between attributes is one of their benefits. As the modulation of gene expression is anticipated to be coupled, it is definitely a non-negligible benefit when researchers are able to infer GRNs. Moreover, they can handle high-dimension [40] scenarios that could make inference complexity in GRNs because their computational cost is at most linear in the feature count. These methods, which are founded on feature selection with tree-based ensemble approaches and are employed in GRNs, are straightforward and general, making them adaptable to many kinds of genomic data and interactions. In fact, tree-based methods have appealing properties [41] and have been applied successfully in the inference of GRNs [42]. Tree-based ensemble approaches have the ability to capture high-order conditional relationships between expression patterns since they are not required to make any assumptions about the nature of the regulation of gene expression. Additionally, because the final model is derived by merging the results from a number of non-linear learners, such practice enables the delivery of good performance with a

DOI: 10.54254/2755-2721/13/20230738

reasonable sample size requirement. Scalability is another benefit of the tree-based technique, some academics have used the non-parametric tree-based method to guarantee the scalability of the entire process [40]. On the other hand, these methods have the same shortcomings as the traditional tree-based approaches. Due to the likelihood of overfitting in singular decision tree models, the predictions are typically not strong. Furthermore, since even a small change in the input dataset can have a significant impact on the outcomes, the models may appear unstable. The practical and famous methods of the tree-based models are GENIE3, GRNBoost, GRNBoost2, and Jump3, which are introduced briefly in the following paragraph.

The predictions of a regulatory network involving p genes are broken down by GENIE3 into p distinct regression issues. Using tree-based ensemble methods from Random Forests or Extra-Trees, the target gene's expression pattern is deduced from the expression patterns of the input genes for each regression problem. It is assumed that a putative regulatory connection is present when an input gene has a significant role in the forecast of the target gene expression pattern. The entire network is then recreated by averaging putative regulatory linkages across all genes to offer a rating of interactions [42]. As a result, GENIE3 generates directed GRNs and enables feedback loops to exist in the network naturally. While waiting, it is quick and scalable. Since the ensemble model is quite robust to the noise from individual decision trees, the researchers don't have to prune the random forest in general. And the larger the number of trees k, the better the performance of the random forest, while large computational cost for large k.Because the GENIE3 are simply based on coexpression, they can contain a lot of indirect targets and false positives.

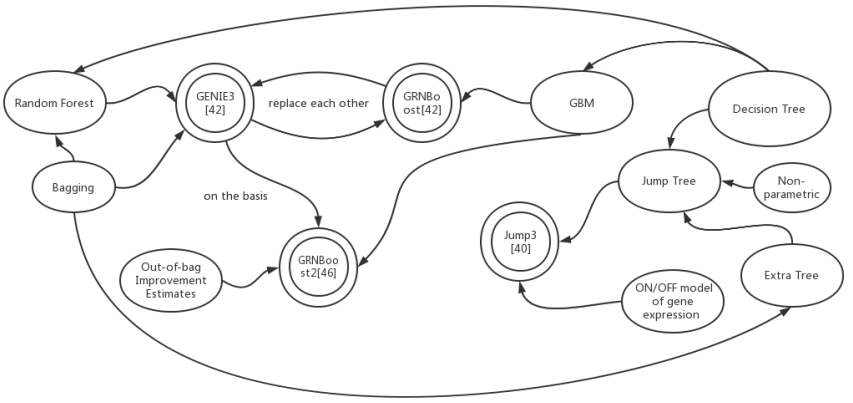

Figure 1. Diagram showing relationships between tree-based algorithms. The double line circle represents the actual method or frame, the ellipse represents the application principle, and the arrow direction represents the direction of deduction.

GRNBoost is based on the notion of suggesting regulators for each target sequence only from the gene expression matrix, much like GENIE3. Nevertheless, GRNBoost utilizes the XGBoost library's version of the gradient-boosting machines (GBM) [43,44], an ensemble learning method that employs boosting [45] as a tactic to merge numerous weak learners, such as shallow trees, into a powerful one. And the base learner in GRNBoost is a regression tree of depth 1 called stumps. Different from the Bagging which GENIE3 uses, Boosting considers the mistakes of previous predictors and trains the new predictors on those mistakes, and then repeats the process till we get a better fit.

GRNBoost2 is a member of the group of regression-based GRN inference techniques, much like GENIE3 [46]. A regularized stochastic variation on GBMs is used in GRNBoost2. Utilizing out-of-bag

DOI: 10.54254/2755-2721/13/20230738

improvement estimates, it provides GBM regression models with a heuristic early-stopping regularization approach [47]. Each regression model then has "just enough" trees in relation to the learning rate. Moreover, regressions that fail to show early net improvement are stopped, preventing unnecessary computational workload. Each new decision tree is trained on a random subset of data, and the remaining values are employed to estimate the gain in the loss function that would come from adding that tree to the ensemble. When the ensemble's early-stopping requirement is satisfied and the average of the last n improvement values falls below 0, no additional trees are created [48].

Jump3, which makes advantage of time series expression data, is based on an official on/off model of gene expression but employs a non-parametric method based on decision trees to rebuild the GRN topologies, allowing the inference of networks containing hundreds of genes [40]. The developers of Jump3 introduce the jump tree approach, which uses the marginal probability of the network node dynamical model as a decision mechanism. As it incorporates the tree-based learning process into the probabilistic model, it successfully establishes itself as a greedy approach to structure learning in a sizable latent-variable model. On the other hand, because this method makes use of the marginal probability, it inherits how simple it is to include side information in probabilistic models. Moreover, Jump3 performs well in terms of scaling, producing results that are comparable to or even superior to those of state-of-the-art competitors on both synthetic and real data.

GENIE3 and GRNBoost2 are immune to any errors in pseudo-time computation since both of them do not require pseudo-time-ordered cells. Moreover, GENIE3 and GRNBoost2 infer directional edges, while they also have multithreaded implementations now. GRNBoost2 was less sensitive to the existence of dropouts, while GENIE3 was more stable throughout repeated runs. Since the bias-reducing impact of gradient boosting allows to employ of shallower decision trees than random forest and employing early stopping, GRNBoost2 generated almost 80% fewer decision trees than GENIE3 in certain circumstances, it performs faster speedups on scRNA-seq datasets than GENIE3 [48].

As mentioned earlier, random forest is the main tree-based algorithm the GENIE3 uses. It is a popular ensemble learner with the approach of Bagging [49], which combines individual weak learners to build strong learners to reduce the variance of the model. Jump3's design is comparable to GENIE3 in that it employs the Extra-Trees technique [50], which randomizes the tests at each divide node of a tree. The Extra-Tree is an improvement over Bagging. It creates a meta-estimator to fit several random decision trees of different subsamples. The Extra-Tree then calculates the average prediction between subsamples. This improves the accuracy of the model and controls overfitting.

As for the applications of the tree-based model in GRNs, SCENIC [51] is a representative one. SCENIC offers vital biological understandings of the factors influencing cellular heterogeneity. The main contribution is to decrease the excessive noise and sparsity of scRNA-seq data by correlating cis-regulatory sequences to the single-cell expression of genes. There are three steps of SCENIC. The first step is to find groups of genes that coexpress with TFs using GENIE3 or GRNBoost. Then, RcisTarget performs cis-regulatory motif analysis in each coexpression module to pinpoint a potential direct-binding target, maintaining considerable pattern enrichment of the right upstream regulator. AUCell generates a binarized activity matrix with decreased dimensionality by scoring the activity of each regulon in each cell. This matrix can be helpful for further analysis in the final phase. Importantly, the GENIE3 and GRNBoost play an essential role in the SCEINC as famous and effective tree-based methods. And in the first stage, the designers swap out GENIE3 for GRNBoost, which is built on Spark [52], to offer a scalable alternative to creating the coexpression network on larger data sets.

3.2. Other Methods

Aside from tree-based methods, which are prevalent for GRN reconstruction, we also summarized other machine learning or deep learning approaches, some of which show better performance compared to tree-based models.

To handle the noise problem of data, Schuldt et al proposed a support vector machine (SVM)-based hybrid approach to learning gene regulatory networks [53]. In the first step of this experiment, developers evaluate the performance of SVMs with different kernel functions. By comparing the

DOI: 10.54254/2755-2721/13/20230738

performance, the developers fixed the SVM kernel to the linear kernel for the next experiment. Thus, in the final experimental results, this method outperformed the previously mentioned Inferelator.In fact, SVM outperforms some deep learning programs because when the working set is a binary classification of a relatively small data set, where the linear relationship between the learned feature vectors and labels is effective for test data classification.

Compared with the classical algorithms GENIE3 and JUMP3 previously described, SINCERITIES has lower computational complexity and can cope with very large dimensional problems while performing better prediction accuracy for GRNs [54]. A major core of this method is to restore the directed regulatory relationship between genes through regularized linear regression. Briefly, the model developers reasoned about the linear regression problem on GRNs after calculating the distribution distance of the time-stamped gene expression data. The linear regression equation of the model for each gene takes into account the distance matrix and solution vector corresponding to the time window corresponding to all genes.

Combining the Generalized Extended Kalman filter (GEKF) with the weight update function in the training algorithm [55], a GRN hybrid model based on recurrent neural network (RNN) is proposed to improve the precision of GRN construction [56]. Biological proximity and mathematical flexibility can be reconciled more effectively with the help of RNN, a complicated neural network that can also capture the intricate, non-linear, and complex interplay between GRN variables. In the RNN-based GRN model, a gene node can calculate the expression at each moment from the expression of all genes associated with that gene and the weights representing regulatory effects. Therefore, the regulatory effect of a particular gene is considered to be the weighted sum of all regulatory genes. The model was tested on the commonly used DREAM Challenge synthesis network and was shown a SOTA performance compared to other approaches [57].

4. Conclusion

As this study suggests, the reconstruction of GRNs is important for us to understand cellular behaviors, and the progression of bioinformatics greatly advanced GRN research. Compared with previous methods such as manual experiments or manual annotation, computational methods have significant advantages, such as better accuracy, easier access to required data, and higher efficiency.

In this article, we will first introduce the traditional methods of GRN, including Boolean networks, ODE, Bayesian networks, and mutual information. We selected some algorithms or frameworks that represent each method, such as ARACNE based on mutual information methods and Inferelator based on ODEs for a brief introduction. With the development of computer technology, the application of machine learning shows better performance than traditional methods in GRN research. On the basis of the use of traditional methods to improve manual experimentation, machine learning can deal with or optimize the shortcomings of traditional methods, such as the inability to deal with highly complex networks. In the introduction of machine learning methods, we focus on the tree-based method because we found that most of the existing efficient and commonly used algorithms and frameworks for reconstructing GRNs use tree-based methods such as random forests. We also briefly introduced three other algorithms or frameworks based on machine learning, including SVM, linear regression, and RNN. And these three methods can show better performance than tree-based methods in some cases.

Different approaches show different advantages and limitations. For traditional methods, mutual information-based methods can process low-expression data with low computational complexity, but these methods cannot account for the interaction of multiple genes. Boolean networks can solve the problem of multiple gene interactions, and the model based on this method is simple and easy to calculate. However, as with mutual information-based methods, most of them can only be used in static GRN research.ODEs-based methods and Bayesian network-based methods can build dynamic networks. However, ODEs-based methods cannot handle large GRN modeling and perform better in linear models than with complex nonlinear adjustment processes. At the same time, the number of network topologies of Bayesian networks increases exponentially with the number of genes, and dynamic Bayesian networks have the disadvantages of increasing computational complexity and

prolonging the time required. Like the decision tree-based methods and other machine learning methods described in this paper, machine learning methods can solve such problems of high dimensionality, high computational complexity, and accuracy. Among them, we found that the tree-based approach solves these problems well and is more efficient. However, the combination of different data types can hurt overall performance. And when multi-model integration occurs, there may be collinearity problems, leading to confounding factors in inference. Other machine learning methods exist in large-scale networks, and the complexity is growing, and the predicted performance is declining. And there are also noise problems.

For research to reconstruct GRNs, both the practical application of GRN and the available data affect its effectiveness. Also, diverse data kinds and objectives give GRN inference approaches their own set of benefits and drawbacks. Of course, it is now popular to reconstruct GRNs with machine learning, and existing machine learning-based methods perform very well. Now researchers are optimizing by combining multiple methods, and combining multiple methods can indeed be more efficient in reconstructing GRNs and improving performance than developing completely fresh methods. However, if multiple methods are simply combined, it will inevitably lead to problems such as the curse of dimensionality. It can also solve the difficulty of obtaining data and bioethical issues such as human gene expression experiments. In summary, reconstructing GRNs with computational approaches is undoubtedly a powerful way to understand complex biological systems. With the improvement of sequencing technology and the innovation of computational techniques, we will be able to learn GRN more comprehensively and accurately in the future.

Acknowledgments

This paper is supported by National Natural Science Foundation of China under Grant Nos. 61502198, 61472161, 61402195, 61103091, U19A2061;the Science and Technology Development Plan of Jilin Province under Grant No. 20210101414JC, 20160520099JH , 20190302117GX, 20180101334JC, 2019C053-3; Research Topic of Higher Education Teaching Reform in Jilin Province under Grant No. 20213F2QZ6100FV; Jilin University Undergraduate Teaching Reform Research Project under Grant No. 2021XYB125

Contact author: Haiyang Jia, jiahy@jlu.edu.cn, Jilin University, 2699 Ave. Qianjin, Changchun, 130012, P.R.China.

References

[1] Latchman DS (September 1996). "Inhibitory transcription factors". The International Journal of Biochemistry&CellBiology. 28 (9): 965–974. doi:10.1016/1357-2725(96)00039-8. PMID 8930119.

[2] Liang L, Gao L, Zou X-P, Huang M-L, Chen G, Li J-J, et al. Diagnostic significance and potential function of miR-338-5p in hepatocellular carcinoma: a bioinformatics study with microarray and RNA sequencing data. Mol Med Rep 2018;17:2297–312.

[3] T.I. Lee, R.A. Young, Transcriptional regulation and its Misregulation in disease, Cell. 152 (2013) 1237–1251, https://doi.org/10.1016/j.cell.2013.02.014.

[4] B. Usadel, T. Obayashi, M. Mutwil, F.M. Giorgi, G.W. Bassel, M. Tanimoto, A. Chow, D. Steinhauser, S. Persson, N.J. Provart, Co-expression tools for plant biology: opportunities for hypothesis generation and caveats, Plant Cell Environ.32 (2009) 1633–1651, https://doi.org/10.1111/j.1365-3040.2009.02040.x.

[5] R. Sibout, S. Proost, B.O. Hansen, N. Vaid, F.M. Giorgi, S. Ho-Yue-Kuang, F. Legée,L. Cézart, O. Bouchabké-Coussa, C. Soulhat, N. Provart, A. Pasha, P. Le Bris,D. Roujol, H. Hofte, E. Jamet, C. Lapierre, S. Persson, M. Mutwil, Expression atlas, and comparative coexpression network analyses reveal important genes involved in the formation of lignified cell wall in Brachypodium distachyon, New Phytol.215 (2017) 1009–1025, https://doi.org/10.1111/nph.14635.

[6] F. Emmert-Streib, M. Dehmer, B. Haibe-Kains, Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks, Front Cell Dev Biol. 2 (2014) 38, https://doi.org/10.3389/fcell.2014.00038.
[7] A.J. Singh, S.A. Ramsey, T.M. Filtz, C. Kioussi, Differential gene regulatory networks in development and disease, Cell. Mol. Life Sci. 75 (2018) 1013–1025, https://doi.org/10.1007/s00018-017-2679-6.
[8] D. Mercatelli, F. Ray, F.M. Giorgi, Pan-Cancer and Single-Cell Modeling ofGenomic Alterations Through Gene Expression, Front. Genet. 10 (2019). doi: https://doi.org/10.3389/fgene.2019.00671.
[9] Fernando M. Delgado, Francisco Gómez-Vela.Computational methods for Gene Regulatory Networks reconstruction and analysis: A review. Artificial Intelligence in Medicine: Volume 95, April 2019, Pages 133-145
[10] Ament, S.A. et al. (2018) Transcriptional regulatory networks underlying gene expression changes in Huntington's disease. Mol. Syst. Biol., 14, e7435.
[11] Singh, A.J. et al. (2018) Differential gene regulatory networks in development and disease. Cell. Mol. Life Sci., 75, 1013–1025.
[12] Margolin, A. A. et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics 7 (Suppl. 1), S7 (2006).
[13] Raza, Khalid, Alam, Mansaf, Recurrent Neural Network Based Hybrid Model for Reconstructing Gene Regulatory Network.Computational Biology and Chemistry http://dx.doi.org/10.1016/j.compbiolchem.2016.08.002
[14] Paolo Mignone, Gianvito Pio, Sašo Džeroski & Michelangelo Ceci. Exploiting transfer learning for the reconstruction of the human gene regulatory network Bioinformatics, 36(5), 15531561(2020)
[15] Buermans H, Den Dunnen J. Next generation sequencing technology: advances and applications. Biochim Biophys Acta (BBA)-Mol Basis Dis 2014;1842:1932–41.
[16] Li Y, Liu L, Bai X, Cai H, Ji W, Guo D, et al. Comparative study of discretization methods of microarray data for inferring transcriptional regulatory networks. BMC
[17] Monger C, Kelly PS, Gallagher C, Clynes M, Barron N, Clarke C. Towards next generation CHO cell biology: bioinformatics methods for RNA-Seq-based expression profiling. Biotechnol J 2015;10:950–66.
[18] Bioinform 2010;11:520.Park, P. ChIP–seq: advantages and challenges of a maturing technology. Nat Rev Genet 10, 669–680 (2009). https://doi.org/10.1038/nrg2641
[19] M.E. Csete, J.C. Doyle, Reverse engineering of biological complexity, Science 295 (2002) 1664–1669.
[20] Mitchell, Tom (1997). Machine Learning. New York: McGraw Hill. ISBN 0-07-042807-7. OCLC 36417892. Archived from the original on 2020-04-07. Retrieved 2020-04-09
[21] Koza, John R.; Bennett, Forrest H.; Andre, David; Keane, Martin A. (1996). "Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming". Artificial Intelligence in Design '96. Artificial Intelligence in Design '96. Springer, Dordrecht. pp. 151–170.
[22] Shrivastava, H.; Zhang, X.; Song, L.; Aluru, S. GRNUlar: A Deep Learning Framework for Recovering Single-Cell Gene Regulatory Networks. J. Comput. Biol. 2022, 29, 27–44.
[23] Sverchkov Y, Craven M. A review of active learning approaches to experimental design for uncovering biological networks. PLoS Comput Biol 2017;13:e1005466.
[24] Barman S, Kwon Y K. A novel mutual information-based Boolean network inference method from time-series gene expression data[J]. PloS one, 2017, 12(2): e0171097.
[25] J. Cao, X. Qi, H. Zhao, Modeling gene regulation networks using ordinary differential equations, Methods Mol. Biol. 802 (2012) 185–197.
[26] A. Madar, A. Greenfield, H. Ostrer, E. Vanden-Eijnden, R. Bonneau, The inferelator 2.0: A scalable framework for reconstruction of dynamic regulatory network models, in 2009 Annual

International Conference of the IEEE Engineering in Medicine and Biology Society, 2009: pp. 5448–5451. doi: https://doi.org/10.1109/IEMBS.2009.5334018.

[27] Bonneau, R., Reiss, D.J., Shannon, P., Facciotti, M., Hood, L., Baliga, N.S. and Thorsson, V. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo, Genome Biol 7:R36 (2006).

[28] B. Usadel, T. Obayashi, M. Mutwil, F.M. Giorgi, G.W. Bassel, M. Tanimoto, A. Chow, D. Steinhauser, S. Persson, N.J. Provart, Co-expression tools for plant biology: opportunities for hypothesis generation and caveats, Plant Cell Environ.

[29] Larjo A, Shmulevich I, Lähdesmäki H. Structure learning for Bayesian networks as models of biological networks. Data mining for systems biology. Springer; 2013. p.35–45.

[30] Deeter A, Dalman M, Haddad J, Duan Z-H. Inferring gene and protein interactions using PubMed citations and consensus Bayesian networks. PLoS One 2017;12:e0186004.

[31] C.E. Manfredotti, Modeling and inference with relational dynamic bayesian networks, Lect. Notes Comput. Sci. 5549 (2009) 287–290.

[32] N.X. Vinh, M. Chetty, R. Coppel, P.P. Wangikar, Gene regulatory network modeling via global optimization of high order Dynamic Bayesian Networks, BMC Bioinf. 27 (2012) 2765–2766.

[33] B.C. Ross, Mutual information between discrete and continuous data sets, PLoS One 9 (2014) e87357, https://doi.org/10.1371/journal.pone.0087357.

[34] Mercatelli D, Scalambra L, Triboli L, Ray F, Giorgi FM. Gene regulatory network inference resources: A practical overview. Biochim Biophys Acta Gene Regul Mech. 2020 Jun;1863(6):194430. doi: 10.1016/j.bbagrm.2019.194430. Epub 2019 Oct 31. PMID: 31678629.

[35] Liang K C, Wang X. Gene Regulatory Network Reconstruction Using Conditional Mutual Information[J]. EURASIP Journal on Bioinformatics and Systems Biology, 2007, 1(2008): 1-14.

[36] Raza, Khalid, Alam, Mansaf, Recurrent Neural Network Based Hybrid Model for Reconstructing Gene Regulatory Network.Computational Biology and Chemistry http://dx.doi.org/10.1016/j.compbiolchem.2016.08.002

[37] W. Zhao, E. Serpedin, and E. R. Dougherty, “Inferring the structure of genetic regulatory networks using information theoretic tools,” in Proceedings of IEEE/NLM Life Science Systems and Applications Workshop (LSSA ’06), pp. 1–2, Bethesda, Md, USA, July 2006.

[38] W. Zhao, E. Serpedin, and E. R. Dougherty, “Inferring connectivity of genetic regulatory networks using information theoretic criteria,” IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 5, no. 2, pp. 262–274, 2008.

[39] Pataskar A, Tiwari VK. Computational challenges in modeling gene regulatory events. Transcription 2016;7:188–95.

[40] Huynh-Thu, V. A. & Sanguinetti, G. Combining tree-based and dynamical systems for the inference of gene regulatory networks. Bioinformatics 31, 1614–1622 (2015).

[41] Geurts, P. et al. (2009) Supervised learning with decision tree-based methods in computational and systems biology. Mol. BioSyst., 5, 1593–1605.

[42] Huynh-Thu, V.A. et al. (2010) Inferring regulatory networks from expression data using tree-based methods. PLoS One, 5, e12776.

[43] Chen, T. & Guestrin, C. In Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 785–794 (ACM,2016).

[44] Friedman, J.H. (2001) Greedy function approximation: a gradient boosting machine. Ann. Stat.,1189–1232.

[45] Freund, Y. & Schapire, R.E. Jinko Chino Gakkaishi 14, 771–780 (1999).

[46] Huynh-Thu,V.A. and Sanguinetti,G. (2018) Gene regulatory network inference: an introductory survey. arXiv: 1801.04087

[47] Sławek,J. and Arodz,T. (2013) ENNET: inferring large gene regulatory networks from expression data using gradient boosting. BMC Syst. Biol., 7, 106.

DOI: 10.54254/2755-2721/13/20230738

[48] Moerman, T. et al. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. Bioinformatics 35, 2159–2161 (2018).
[49] Breiman, L. Bagging predictors. Mach Learn 24, 123–140 (1996). https://doi.org/10.1007/BF00058655
[50] Geurts, P. et al. (2006) Extremely randomized trees. Mach. Learn., 36, 3–42.
[51] Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine J-C, Geurts P, Aerts J et al (2017) SCENIC: single-cell regulatory network inference and clustering. Nat Methods 14: 1083–1086
[52] Zaharia, M. et al. In Proc. of the 9th USENIX Conference on Networked Systems Design and Implementation 2–2 (USENIX Association, 2012).
[53] Schuldt C, Laptev I, Caputo B. Recognizing human actions: a local SVM approach[C]//Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. IEEE, 2004, 3: 32-36.
[54] Papili Gao N, Ud-Dean S M M, Gandrillon O, et al. SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles[J]. Bioinformatics, 2018, 34(2): 258-266.
[55] Julier, S.J., & Uhlmann, J.K.;1; (1997). New extension of the Kalman filter to nonlinear systems. In AeroSense'97 (pp. 182-193). International Society for Optics and Photonics.
[56] Raza K, Alam M. Recurrent neural network based hybrid model for reconstructing gene regulatory network[J]. Computational biology and chemistry, 2016, 64: 322-334.
[57] Stolovitzky G, Monroe D, Califano A.;1; Dialogue on reverse-engineering assessment and methods: the DREAM of high-throughput pathway inference.Annals of the New York Academy of Sciences.2007;1115(1):1-22.