THINK INDIA JOURNAL

ISSN:0971-1260

Vol-22-Issue-41-December-2019

Positive Aspects for Reconstruction of Gene Regulatory Network

Sandeep Mittal[1], Abhishek Bhola[2] and Hariom Tyagi[3]
[1,3]Department of CSE, R D Engineering College
[2]School of ICT, Gautam Bhuddha University, Greater Noida

ABSTRACT

In the current years, the accumulation of Genetic Data has increased and itencouragesto formulatesome novelmethods for efficient handling of geneticinformation. This huge amount of genetic informationpresent in various data repositories can helps in studying the complex biological processes of living organisms. Gene Regulatory Network has proved to be the most promising means for analyzing the biological processes. In Gene Regulatory Network, genes act as the nodes and edges show relationship between them andanalysis of these networks eventually guides us to discover new drugs for human beings and helps to understand various biological processes at genetic level.Reconstruction refers to the method of network inference out of experimental results.In this paper we have summarized the positive aspects of GRNs and also carried out the comparative analysis of significant GRN reconstruction methods presented in the literature.

Keywords: Gene Regulatory Network, Bioinformatics, Bayesian Network, Rough Set Theory, Microarray data, Knowledge Database Discovery, Differential Equations, Boolean Networks, Artificial Neural Networks, Fuzzy Model, Microarray Data

INTRODUCTION

The concepts discuss in this paper comes under the field of Bioinformatics. Bioinformatics is thefield of studyin which biology, computer science and information technology are merged into a sole stream in other words, itcan also be understood as the field where computer methods are utilized to store, manage, maintain and analyzebiological datasets. Genes regulates all the biological processes of a living organism and synthesize proteins that are essential for building and working of a living organism.

Therefore,genes and their interactions with each other are vital and thus form a biological network called as gene regulatory network (GRN). In gene regulatory network,genes act as nodes and edge between two genes show how one gene regulates other in positive/negative way affecting the production of various proteins.

Reconstruction of gene regulatory network means constructing itwith the help of available biological datasets such gene expression data, DNA/RNA sequencing data, etc.

Themotivation to explore this area of study is due to its application in disease diagnoses, prevention, drug designing, etc. It also helps in understanding the complex process of evolution of life of living organisms. In the next section 2, the literature review of some important work carried out in the field have been summarized with their comparative analysis.In Section 3, various challenges faced by various researchers, scientists and bioinformatician are also listed which is finally followed by the conclusion section.

LITERATURE REVIEW

In this section we have epitomized the previous works done in the field of gene regulatory network reconstruction in order to find the challenges and comparative analysis has been also done.

[1] Sudip Mandal et al. have reconstructed GRNs from microarray data by making use of Rough Set Theory and Bayesian approach.Microarray datais utilized in the medical domain torepresent Genetic profile of unhealthyand normal tissues of patients. Rough Set Theory is capable of extracting various dominant Genes that portray a major part in causing the disease and also helps to produce a uniquesimplified rule set for creating expert systems in medical sciences with high precision. RST works well to find dependence relationship among Microarray data. Now after RST, Bayesian Network that uses Direct Acyclic Edgesto

show relationship between nodes Each node i.e. a Gene is attached with other Gene with a mutual probability distribution table to show dependency relationship between nodes. This approach is probabilistic in nature, is having capability of handling the noise and estimating the confidence in the differentfeatures of the network. Accuracy of the mentioned approach is not up to the mark because of low quality ofdata collected by microarray technology, also Gene Regulatory Network is constrained by limited number ofsamples(patients).

We want to eliminate the limitation of GRN, as we know GRN has stated itself an important tool for the analysis of interaction among different Genes. Further this analysis can help in understanding various complex biological processes. In this sense research on GRN must continue, so following this idea various inference methods have been suggested by[2] Fernando M. Delgado et al. Inference Method of GRN having a process that depends on renowned Knowledge Database Discovery workflow i.e. KDD. KDD is a process of discovering useful information from the collection of Data. In case of GRN, it passes through various stages i.e. pre-processing of the input data to validating the formulated models of GRN. Pre-processing of data before inferring GRN, is a crucial step for reconstructing GRN and it also enhance the quality of outcome.

The important inference methods that are summarized by Fernando M. Delgado et al.are (a) Information Theory (b) Boolean Networks (c) Differential Equations (d) Bayesian network methods.Information Theory models are the simplest modelsand effective also because they cover different aspect of cells. They are capable of discovering large GRNs from limited number of samples. Boolean Networks that allow two states mainly i.e. true and false, are easy in implementation. Now another Inference method is Differential Equations that mainly utilize continuous variable instead of Discrete variables. Bayesian Networks are the most used GRN inference architectures, they make use of Bayes theorem of probability, the combining probability and graph theory to qualitatively model the properties of GRNs.Bayesian Networks are the most prominent methods of reconstructing GRN that use theorem of probability and Graph theory to model the GRN but they could not handle large networks, so in further section we going to discuss other methods.

[3] Ray Duenas et al. implements the method that consists of applying a unique one genetic algorithm for eachgene Independently, instead of applying a unique genetic algorithm to determine the whole network. To guide the Genetic algorithms this method has used Akaike Information criterion as fitness function, further Boolean NetworksConfirm their accuracy. Boolean Networks are important where samples are limited in numbers. GRN Inference methods based on genetic algorithm, where it is applied one genetic algorithm per Gene independently with the purpose of obtaining best predictor genes for each target gene. This method is useful because it can operate wheresamples are limited in numbers, also in this method each gene is given individual importance and it helps in enhancing the accuracy of result i.e. further confirmed by the Boolean Networks. These Networks are easy in interpretation but this method has high computing time.

[4] GaoYanpu et al. discussed a new method thattransforms a differential equation model of a Gene Regulatory Network into a discrete equation by using explicit Euler method. Differential equation modelsare continuous Models that are considered to be suitable for describing the continuous gene dynamics. Aleast square squares algorithm is represented here to generate parameters estimates by replacing the unknowntrue outputs with their estimates. Further simulation results indicate that the proposed algorithm is effective and is of a high estimated accuracy. The model parameters estimation methods further can be divided intothe recursive methods, the iterative methods, the over parameterization methods and the hierarchical identification methods.

As we are encountering various computational methods one by one, [5] Ed Keedwell et al. are going todiscuss a new method named Artificial Neural Networks used for constructing GRN. This method is used to assimilate the microarray data and constructs Gene Regulatory Networks. Microarray data contains experimental noise,has floating nature and ANNs (artificial neural network) in purest form is not capable of dealing withthese problems, so ANNs is combined with generic algorithm and genetic data is collected for making GRNBefore this, standard neural networks have not been used. Further Boolean Networks confirm the validity of the ANNs responses with the networks which are used to create the data. Gene Regulatory Networks help us to study and inspect genes role in biological processes so they have utmost importance and with advent of time many inference methods have been invented. In this paperGRN is reconstructing by using temporal microarray data using Pattern Recognition Technique. The data set is created here by the help of a simulation model. For Data Mining they made use of order estimation criteria combining with artificial neural networks. The mentioned approachis useful in inferring meaningful genetic data from large data set, it allows to associate priori-biological knowledge and thus lessens the dimensionality. So,at the end this approach generates more decisive and precise models of GRNs

[6]Azhar Salimet al. discuss reconstructing GRNs by the help of temporal microarray data using Pattern Recognition Technique. The data set is created here by making use of asimulation model. For DataMining they utilize order estimation criteria, in union with artificial neural networks. This approachis useful in inferring meaningful genetic data from large data set, it allows to incorporate priori-biologicalknowledge and thus reduces the dimensionality. So, at the end this approach produces more decisive and error-less models of GRNs.

[7] Young Hwang Chang et al. discuss a method that infersGene Regulatory based on Repairing Sparse Rank Sparse matrices. This Method is quite useful because most of the methods concentrate on making Gene Regulatory Networks but a but a little concentration is given to the fact that How all these networks operate under different environment conditions. We face challenges not only inmaking GRNs but also how these networks conclude results with different parameters. This concept eliminates this problem by giving a method which enables us to estimate bias errors due to effectof perturbations and correctly identify the common graph structure among biased inferred graph structures.To do this task, common dynamics of the GRN subject to various perturbations is retrieved. The task of removing the errors has been referred to “repairing” in this concept. The method we already mentionedposture the problem as sparse Low-rank representation problem, by contriving the network inference as finding a sparsely connected structure that has low rank over multiple experiments. They design a unique optimization formulation which enables us to combine temporal data sets from various perturbations experiments. The method can automatically repair the common graph structure from the data sets of perturbed GRNs, even without having precise information about the effect of perturbations.As we are encountering various computational methods for evaluating GRN, here we want to highlight another Computational method. The arrival of microarray has given task to many Computer Scientists to evolve better inferring methods to represent relationship in between the Genes.

[8] Ramesh Ram et al. proposes a modelthat has been intended to find microarray datasets for regulatory relationships in between the genes. This model is refined for finding regulatory triplets by the help of predicting changes in expression level, and comparing them with real changes. This model eradicates false predictions from the classical fuzzy modeland introduces wider space for inferring regulatory relationships. There also has been introduced a unique pre-processing technique that uses Fuzzy logic that can congregate genes having similar changes in expression profileat all intervals in the microarray data. They practice all possible combination of activators and repressorsfor each particular gene and the output given by the model is compared to the expression level of the rest of the Genes. The various unions of genes are given ranks on the basis of error between anticipated and target gene. The unions that have less error, are taken into consideration.

[9] Frank Emmert-Streib et al. discusses the importance of GRNs and its applications. An effort has been put to bring GRNs closer to the clinical and medical applications. Inference of GRNs is a challenging task, because it goesbeyond the mere association between such entities that would comprise also indirect relations involving intermediate gene products as well. When number of GRNs from different physiological and diseased conditions becomeavailable, it will be likely to compare these networks. This will allow to understand about the interaction changes across different physiological or disease conditions and supplement our biologicaland biomedical understanding. Further that knowledge helps us in understanding the complex process of living cell of organisms, inventing new drugs and other important tasks related to medical applications.

The gene regulatory network reconstruction method discussed above have been also compared on the basis of different parameters such as technique used, parameters, advantages and disadvantages is shown in Table 1.

Table 1 Comparative Analysis of various Gene regulatory Network reconstruction methods

References	Technique Used	Parameters	Advantages	Disadvantages
1.Sudip Mandal et al.	Rough Set Theory, Bayesian Network	1.RST finds set of dominant Genes. 2.Direct Acyclic Edges and Graph Theory- each Gene joined with other with probability distribution	1. It is quite flexible 2. It does not require a large number of variables	1. Low accuracy 2. GRN is constrained due to limited number of samples
2. Ray Duenas et al.	Bayesian Network, Akaike Information Criterion and Boolean Network	1.Akaike Information Criterion- that helps in applying algorithm for each Gene independently 2. Bayesian Network installs the graph among various Genes. 3.Boolean Networks confirm accuracy of that method.	1.This method is very useful where we have limited number of samples. 2.It has high accuracy because it deals with each Gene individually	1.It is efficient only in limited number of Genes. 2.This method is not suitable for large networks and hence large networks cannot be inferred with the mentioned method
3.Rao Yanpu et al.	Euler Method, Differential Equation	Euler method transforms the continuous Differential Equation of a GRN into Discrete Differential Equation.	1.Differentail Equations work on the continuous data while this method transforms a Differential Equation of GRN into Discrete Equation using Euler method. 2.This method is good in terms of accuracy.	1.This method is complex in application.
4.Ed Keedwell et al.	Microarray Data, Artificial Neural Network and Boolean Network	1.Artificial Neural Network not in purest form installing GRN from Microarray Data 2.Boolean Network further tests the validity of Network generated by ANNs.	1.This method is capable of handling noise present in the data. 2.Manage non-linear and dynamic behavior.	1.Due to Computational complexity, this method is suitable for very small systems.
5.Azhar Salim et al.	Pattern Recognition Technique	1.Order Estimation Criteria- This is applied to assess the set of Genes that are involved in the regulation of every Gene. 2.Artificial Neural Network- after Order Estimation Criteria ANN finally installs the network	1.It reduces the Curse of Dimensionality 2.This method has high accuracy than other methods.	1.As Order Estimation Criteria involves number of complex models that require specialized knowledge and it is not easy for

		among Genes.		everyone to implement.
6.Young Hwan Chang et al.	Repairing Sparse Low Rank Matrices	1.Formulating GRN as a Dynamical System whose Parameters are time invariant 2.After that the integration of different experiment data and deriving a sparse and Low rank Matrix 3.Now formulation of the Network after finding a sparsely connected structure that has low rank over various experiments.	1.This method takes care of the effect of perturbations on GRN. 2.It can automatically repair the Common Graph Structures across perturbed GRNs and even without having knowledge about them.	1.This method cannot avoid ambiguity present in the Data.
7.Ramesh Ram et al.	Fuzzy Model	1.This model includes Fuzzy logic that can group genes having similarity in expression profile 2.It tries every possible combination of activators and repressors for each particular gene 3.The Combination of genes are given ranks on the basis of error between targeted and predicted Gene, the combinations that have low error are taken into consideration.	1.This method is consistent 2.It improves accuracy as it includes low error between Targeted and Predicted Gene.	1.This method needs more Computational time as compared to other methods.

Positive aspects mean the Brighter side of studying about GRN, various important reasons that Motivate us to infer or reconstruct GRN. Gene is the basic building block of human life and it also takes part in Protein Synthesis. In that case we are expressing importanceof one Gene then thousands of Genes interacting with each other in GRN cannot be ignored. Number of Genes interacting with each other affects the rate of production ofProtein from every Gene, the process is known as “Gene Regulation”. Further GRN helps inrevealing the inner complex mechanism of adaptability of various living organisms to the environment. GRN also plays an important role in biomarkers for Diseases that further help inDrug Design. Modelling of Gene helps in discovering and portraying powerful dynamics of complexof complex biological systems involved in various cellular activities such as development, ageing etc.GRN can contribute in number of fields such as Genetic, improving Diagnostic procedures, System Dynamics, producing Drug Design with greater impact, System Dynamics etc.

CHALLENGES

There are number of difficulties while reconstructing GRN like estimation of GRNis difficult for large scale Networks because search space grows exponentially with number of Genes, additionally Genes collectedfrom most of the methods are subjected to noise.Further Microarray technique i.e. used in collection of Gene Expression Data at different time points suffers from Dimensionality problemmeans higher number of Genes (in thousands) and a relativelylow number of samples which are in dozens.Now various computational methods that are used in reconstructing GRN, have their own limitations e.g. Bayesian Networks cannot have cycles and also this method is not appropriate for time series data, Dynamic Bayesian

Network is complicated in terms of computations, in Boolean Networks discretization of Data is done which directs to Data loss. Most of the models consider Linear interaction among the Genes while interaction among them is non-linear.

CONCLUSION

In this paper we have discussed importance of GRNs and epitomized various works done in this field.Every work is signifcnt and relevant in itself and it's very difficult to choose which work is significant as each work isconsidering different parameters and constraints. But according to our study and comparative analysis we can say work done by Young Hwan et al. is most significant out all the above because in this work attention is not given not only to themaking of networks but also how these networks operate under different environment conditions and howthe results generated by Networks vary with different parameters. Although other works like by Sudip Mandal et al. is also relevant, result generated by Rough Set Theory and Bayesian method together is betteras compared to other methods. This is clear that Gene Regulatory Networks has utmost importance because in coming years how human live evolves and what types of drugs a human life would need can be inferred bythe help of Gene Regulatory Networks. We can progress in this direction with computational methods havinggood accuracy under different parameters and under different environment conditions and therefore the resultchanges must be considered and analyzed.

REFERENCES

[1] Sudip Mandal, Goutam Saha, Rajat Kr Pal, "Reconstruction of Dominant Gene Regulatory Network from Microarray Data Using Rough Set and Bayesian Approach" J Computer Science System Biology Volume 6(5): 262-270 (2013) – 264 ISSN: 0974-7230 JCSB

[2] Fernando M. Delgado, Francisco Gomez-Vela, "Computational methods for Gene Regulatory Networks reconstruction and analysis" ,(2019) Artificial Intelligence in Medicine, 95 , pp. 133-145.

[3] Ray Duenas, David Correa Martins-Jr, Carlos Silva Santos, "Gene Networks through one Genetic Algorithm per Gene", 2014 IEEE International Conference on Bioinformatics and Bioengineering.

[4] GAO Yanpu, WANG Dongqing, "Least Squares Identification Method for Differential Equations of Gene Regulatory Networks", Proceedings of the 33rd Chinese Control Conference July 28-30, 2014, Nanjing, China

[5] Ed Keedwell, AjitNaraynan and Dragan Savic, "Modelling Gene Regulatory Data using Artificial Neural Networks", Neural Networks, Proceeding of the 2002 International Joint Conference on Neural Networks, V

[6] Azhar Salim and FaramarzValafar, "Reconstruction of Gene Regulatory Networks From Temporal Microarray Data Using Pattern Recognition Technique",Published in: The 2006 IEEE International Joint Conference on Neural Network Proceedings

[7] Young Hwan Chang, Roel Dobbe, Palak Bhushan, Joe W.Gray and Claire J.Tomlin, "Reconstruction of Gene Regulatory Network Based on Repairing Sparse Low Rank Matrices", IEEE/ACM Transactions on Computational Biology and Bioinformatics (Volume: 13 , Issue: 4 , July-Aug. 1 2016)

[8] Ramesh Ram, Madhu Chetty and Trevor I. Dix, "Fuzzy Model for Gene Regulatory Network", Published in: 2006 IEEE International Conference on Evolutionary Computation

[9] Frank Emmert-Streib, Matthias Dehmer and Benjamin Haibe-Kains, "Gene Regulatory Networks and their applications : Understanding Biological and Medical Problems in terms of Networks", Published in: Frontiers in Cell and Development Biology in August 2014, Volume: 2, Article: 38

[10] Abhishek, Shailendra Singh, "A Gene Regulatory Network Prediction Method using Particle Swarm Optimization and Genetic Algorithm", International Journal of Computer Applications (0975 – 8887) Volume83 – No 12, December 2013

[11] Paras Bansal, Shailendra Singh, “A review on Bayesian Network Techniques for inferring Gene Regulatory Networks”, International Journal of Advance Research in Science and Engineering Vol. No.4, Special Issue (01), April 2015

[12] Raviajot Kaur, Abhishek Bhola, Shailendra Singh, “A Novel Fuzzy Logic Based Reverse Engineering Of Gene Regulatory Network”, Future Computing and Informatics Journal, Volume 2, Issue 2, December 2017.

[13] Abhishek Bhola, Shailendra Singh, “Gene Selection Using High Dimensional Gene Expression Data: An Appraisal”, Current Bioinformatics, Volume 13, Number 3, 2018, pp.225-233(9)