A REVIEW ON BAYESIAN NETWORK TECHNIQUES FOR INFERRING GENE REGULATORY NETWORKS

Paras Bansal[1], Shailendra Singh[2], Abhishek[3]

[1,2,3]*Department of Computer Science & Engineering,*

PEC University of Technology, Chandigarh (India)

ABSTRACT

Gene regulatory networks are gaining lots of importance among researchers now-a-days. Gene regulatory networks (GRN's) play an important role in evolution of every living being. GRN's are the networks of interaction among genes where nodes represent genes and edges represent interaction among them. GRNs are also of great use in drug designing research as more effective drugs can be produced by using precisely predicting Gene regulatory network tools and methods. Researchers have proposed many approaches to infer Gene regulatory networks. Bayesian networks approaches has been widely used due to its stochastic nature. This paper presents a review of Bayesian networks and its extension in Gene regulatory networks.

Keywords : Bayesian Networks, Directed Acyclic Graphs, Gene Regulatory Networks (GRN's)

I. INTRODUCTION

Research in cellular biology has attracted many researchers in last decade. There has been a shift in biological research. The focus of research was study of smaller molecules such as RNA, DNA, and Proteins etc. But now researcher have started thinking about how these molecules are related to each other. Now-a-days researchers believe that the system can be understood as whole rather than to analyze each molecule differently.

Genes play an important role in production of proteins which are building blocks of living being system. Information is required for production of proteins from genes in genome of every cell. This information controls which genes will produce which protein from that cell. This process of production of protein is central dogma of molecular biology. Some proteins regulate production of different proteins via interaction with different genes. Now this interaction of proteins with genes may have different impact on production of proteins. They may enhance the production of proteins and are known as *activators*.

In contrast they may reduce the production of proteins and are known as *suppressors* or *inhibitors*. These produced proteins may interact with other genes for production of other proteins and this cycle goes on. This genomic level interaction can be represented as a network where nodes represent the proteins or genes and edges represent their relationship i.e. activation or inhibition. No edge means two are not related directly. This special network of interaction of genes, proteins is known as Gene Regulatory Networks. At particular experimental conditions status of genes and their effect on other genes is observed to construct GRN's.

Reconstruction of GRN is quite interesting but difficult task for researchers. In medical, root of many diseases caused due to improper functioning of regulatory processes can be simplified by using effective tools for predicting GRN. Many costly and tedious experiments will be replaced by these GRN inferring methods which will save lots of effort, time and money involved in large biotechnological projects. GRN can also be used to

study and analyze the long term effect of drugs in our body. This will also be helpful in analyzing the dynamics of genes under drugged, diseased or experimental conditions. Hereditary diseases can also be analyzed and studied in greater detail with precise GRN inference methods. With these advantages many techniques have been proposed to model and infer gene regulatory networks.

In literature many reviews[1]–[5]have been made which consider all the major approaches. One of most widely used is Bayesian network for construction of Gene regulatory networks. This method considers the expression data as set of random variables and associates probabilities with them. Then these probabilities are used to construct networks. In this paper a review on the available techniques/variations in GRNs using Bayesian networks has been presented. At the outset, the overview of all other approaches and the basics of Bayesian networks are given. Then the brief description of different computational approaches for construction of GRNs is highlighted followed by description of Bayesian network techniques to infer the GRNs. At last the conclusion is given.

II. COMPUTATIONAL APPROACHES TO INFER GRN'S

Gene regulatory networks show the interaction among a number of genes. GRNs should be accurate and reliable as much as possible. Different approaches like Boolean networks [6]–[8], Probabilistic Boolean network[9]–[12], Ordinary Differential Equations [13]–[15], Artificial Neural Networks [16]–[18], Bayesian Networks have been used to maximize the accuracy and reliability of the constructed GRNs. The description of the Bayesian network and its various variations used to infer GRNs is mentioned below

III. BAYESIAN NETWORK TECHNIQUES TO INFER GRNS

Bayesian networks have been widely used in modelling gene regulatory network because of their stochastic nature. These networks are also noise resistant in nature and can also deal with missing gene expression data values. These can also model GRN's quantitative aspects. These network poses some difficulties such as learning is categorized as NP-hard problem for Bayesian networks. But to counteract these problem certain extensions of Bayesian networks have been proposed that use heuristic search methods to search for best network topology having highest score defined according to certain criteria. In this section the brief introduction of the Bayesian networks is given followed by some extensions or applications of Bayesian networks to construct gene regulatory networks

3.1 Bayesian Networks

Bayesian networks are directed acyclic graphs (DAGs). DAG structure is represented using tuple D= (Z, E), with a set of local probability distributions C, where Z represents the set of nodes (which represent random variables) and E represents the set of directed edges (which are direct dependencies between random variables). C has local probability distribution for each node Z_i conditioned on parents of that node, represented as p(Z_i|parents(Z_i)). In particular if there is an edge Z_i to Z_j then it represents there is dependency of variable Z_j onZ_i. In other words, value of variable Z_j is dependent of value of variableZ_i. Also, in this case Z_i is referred as parent and Z_j is referred as child.

3.2 Bayesian Network To Handle Perturbed Expression Profiles

Pe'er et al.[19] used the framework given by Friedman[20] to infer finer relations from perturbed gene expression profiles. They try to answer questions which deal with finer structure. For example, which genes are responsible for interaction between cluster of gene or within clusters? What is the type of interaction between genes? Gene expression for each gene was treated as a random variable provided by gene expression. Most standard methods focuses on pairwise relationships of genes but biological process of gene regulatory network may not be so simple, there may be a chain of mediators between a pair of genes. Pe'er et al. also addresses this concept during inference of gene regulatory networks.

3.3 Bayesian Networks Combined Expression Data With Prior Knowledge

Werhli and Husmeier[21] reconstructed gene regulatory networks with the use of prior biological information to obtain better results. Prior knowledge is represented in form of energy functions. They derived and tested Markov Chain Monte Carlo (MCMC) sampling to sample networks and hyper-parameters from posterior distribution. This enabled automatically learning by system to tradeoff between information from expression data and prior biological knowledge. This approach was combined with Bayesian coupling scheme to reconstruct gene regulatory networks from datasets, which were sampled under different experimental conditions.

3.4 Bayesian Networks And Cross Correlation For Large Number of Genes

Yavari et al.[22]tried to reconstruct gene regulatory network for larger number of genes. Increase in number of genes cause number of possible graphs to increase exponentially and this makes an exhaustive search intractable. So number of genes is increased but they are clustered using existing biological knowledge (gene ontology annotations). After applying clustering, Bayesian network is used to model casual relations among genes in individual groups. And the sub-networks are integrated to make a global network. Also, there is time delay in real GRNs but Bayesian network is a static method which does not consider time delays so Yavari et al.[22]proposed a technique to include time delay in Bayesian networks using concept of cross correlation applied on co-clustered genes

3.5 Bayesian Network Based Sparse Graph Search Algorithm

Yang et al.[23]proposed an algorithm based on Bayesian network to reduce computational time of reconstructing gene regulatory networks. The authors called it Sparse Graph Search algorithm (SGS). They used heuristic approaches to find optimal network in network space

3.6 Bayesian Model Averaging Approach To Infer Large Scale Grns

Kim and Gelenbe[24] presented a technique for reconstructing large scale gene regulatory networks. Constructing large scale GRNs for large number of genes has always been a challenging problem. They combined many appropriate models using Bayesian model averaging approach (BMA). This method is applied on human brain tumor dataset and three large scale networks were built and it included 4422 genes

3.7 Bayesian Network With Uniting Of Partial Problems

Watanabe et al.[25] presented an estimation method of inferring gene regulatory network. Partial problems were united using Bayesian networks. Two problems were addressed. This method can estimate large scale networks.

As real gene regulatory system may contain cyclic structure, this method allows cyclic structure in graph. Firstly whole problem is divided into partial problems of three genes (triplet) each. After dividing problem into sub-problems, scores of all possible DAGs for each sub problem is calculated. Next the three-gene networks are combined to solve original problem

IV. CONCLUSION

This paper has discussed many methodologies using Bayesian network framework to infer gene regulatory network. Some of these methods have reduced computational time and some have tried to reduce dimension of dataset using some preprocessing of datasets. Brief overview of these methods and techniques is given in TABLE 1. The problem of dimensionality is major focus of future research. Another problem is to infer precise network for large number of genes. There has been tremendous growth in this area of research in last few decades. Tools have been developed to reduce dimensionality to increase accuracy and precision in GRN's. At the end, there is need for faster and efficient algorithms that can infer GRNs not only precisely but also for large number of genes.

Table 1: Overview of Methodologies/Techniques Using Bayesian Network

Ref. No.	Description	Dataset	Key points
[19]	Inferred network for perturbed gene expression profiles	Saccharomyces cerevisiae	• Discretization process applied to datasets • Inter-cluster interactions also discovered
[21]	Combined the expression data with prior knowledge	Saccharomyces cerevisiae Synthetic RAF-pathway	• a Bayesian coupling scheme for learning gene regulatory networks • better network accuracy
[22]	Clustered the genes and applied cross correlation to include time delay	Saccharomyces cerevisiae	• Accuracy and sensitivity increased
[23]	Sparse Graph algorithm based on Bayesian network	Escherichia coli	• Computational time reduced,detected relations between 100 nodes in few minutes
[24]	BMA approach for large scale networks	Human brain tumor	• Prior knowledge added • Uncertainties in model selection taken into account • Better performance
[25]	Bayesian network with uniting partial problems	Mouse adipocyte and osteoblast	• More accurate with reduced computational time

REFERENCES

[1] H. de Jong, "Modeling and simulation of genetic regulatory systems: a literature review.," *J. Comput. Biol.*, vol. 9, no. 1, pp. 67–103, 2002.

[2] C. Sima, J. Hua, and S. Jung, "Inference of gene regulatory networks using time-series data: a survey.," *Curr. Genomics*, vol. 10, pp. 416–429, 2009.

[3] C. Panse and M. Kshirsagar, "Survey on Modelling Methods Applicable to Gene Regulatory Network," *Int. J. Bioinforma. Biosci.*, vol. 3, no. 3, pp. 13–23, 2013.

[4] N. Vijesh, S. Chakrabarti, and J. Sreekumar, "Modeling of gene regulatory networks: A review," *J. Biomed. Sci. Eng.*, vol. 2013, no. February, pp. 223–231, 2013.

[5] L. E. Chai, S. K. Loh, S. T. Low, M. S. Mohamad, S. Deris, and Z. Zakaria, "A review on the computational approaches for gene regulatory network construction," *Comput. Biol. Med.*, vol. 48, pp. 55–65, 2014.

[6] T. Akutsu, S. Miyano, and S. Kuhara, "Identification of genetic networks from a small number of gene expression patterns under the boolean network model," *Proc. Pacific Symp. Biocomput.*, vol. 28, pp. 17–28, 1999.

[7] S. a Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets.," *J. Theor. Biol.*, vol. 22, pp. 437–467, 1969.

[8] S. Martin, Z. Zhang, A. Martino, and J. L. Faulon, "Boolean dynamics of genetic regulatory networks inferred from microarray time series data," *Bioinformatics*, vol. 23, no. 7, pp. 866–874, 2007.

[9] R. Pal, A. Datta, M. L. Bittner, and E. R. Dougherty, "Intervention in context-sensitive probabilistic Boolean networks," *Bioinformatics*, vol. 21, no. 7, pp. 1211–1218, 2005.

[10] W. K. Ching, S. Zhang, M. K. Ng, and T. Akutsu, "An approximation method for solving the steady-state probability distribution of probabilistic Boolean networks," *Bioinformatics*, vol. 23, no. 12, pp. 1511–1518, 2007.

[11] R. Pal, A. Datta, and E. R. Dougherty, "Optimal infinite-horizon control for probabilistic Boolean networks," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2375–2387, 2006.

[12] J. Liang and J. Han, "Stochastic Boolean networks: An efficient approach to modeling gene regulatory networks," *BMC Syst. Biol.*, vol. 6, no. 1, p. 113, 2012.

[13] J. J. Tyson, A. Csikasz-Nagy, and B. Novak, "The dynamics of cell cycle regulation," *BioEssays*, vol. 24, pp. 1095–1109, 2002.

[14] T. Chen, H. L. He, and G. M. Church, "Modeling gene expression with differential equations.," *Pac. Symp. Biocomput.*, pp. 29–40, 1999.

[15] H. de Jong and M. Page, "Search for steady states of piecewise-linear differential equation models of genetic regulatory networks," *Comput. Biol. Bioinformatics, IEEE/ACM Trans.*, vol. 5, no. 2, pp. 208–222, 2008.

[16] K. Raza, M. Alam, and R. Parveen, "Recurrent Neural Network Based Hybrid Model of Gene Regulatory Network," *ArXiV*, no. i, p. 18, 2014.

[17] N. Noman, L. Palafox, and H. Iba, "Reconstruction of Gene Regulatory Networks from Gene Expression Data Using Decoupled Recurrent Neural Network Model," *Nat. Comput. Beyond SE - 8*, vol. 6, pp. 93–103, 2013.

[18] R. Xu, D. Wunsch Ii, and R. Frank, "Inference of genetic regulatory networks with recurrent neural network models using particle swarm optimization.," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 4, no. 4, pp. 681–92, 2007.

[19] D. Pe'er, a Regev, G. Elidan, and N. Friedman, "Inferring subnetworks from perturbed expression profiles.," *Bioinformatics*, vol. 17 Suppl 1, pp. S215–S224, 2001.

[20] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, "Using Bayesian Networks to Analyze Expression Data," *J. Comput. Biol.*, vol. 7, pp. 601–620, 2000.

[21] W. V Adriano and H. Dirk, "Reconstructing Gene Regulatory Networks with Bayesian Networks by Combining Expression Data with Multiple Sources of Prior Knowledge," *Stat. Appl. Genet. Mol. Biol.*, vol. 6, 2007.

[22] F. Yavari, F. Towhidkhah, and S. Gharibzadeh, "Gene Regulatory Network Modeling using Bayesian Networks and Cross Correlation," *2008 Cairo Int. Biomed. Eng. Conf.*, pp. 0–3, 2008.

[23] B. Yang, J. Zhang, J. Shang, and A. Li, "A Bayesian network based algorithm for gene regulatory network reconstruction," *2011 IEEE Int. Conf. Signal Process. Commun. Comput.*, pp. 1–4, 2011.

[24] H. Kim and E. Gelenbe, "Reconstruction of Large-Scale Gene Regulatory Networks Using Bayesian Model Averaging," *2011 IEEE Int. Conf. Bioinforma. Biomed.*, vol. 11, no. 3, pp. 202–207, 2011.

[25] Y. Watanabe, S. Seno, Y. Takenaka, and H. Matsuda, "An estimation method for inference of gene regulatory net-work using Bayesian network with uniting of partial problems," *BMC Genomics*, vol. 13, no. Suppl 1, p. S12, 2012.