International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622
International Conference on Industrial Automation and Computing (ICIAC- 12-13th April 2014)

RESEARCH ARTICLE OPEN ACCESS

A Bayesian Model Framework for Reconstructing Gene Network

Ms. Sneha S. Borkar *, Prof. Rakhi Wajgi **
*(Department of Computer Technology, YCCE, Nagpur.
** (Department of Computer Technology, YCCE, Nagpur.

ABSTRACT-
Gene regulatory networks provide the systematic view of molecular interactions in a complex living system. Reconstructing large-scale gene regulatory networks is challenging problems in systems biology. For reliable gene regulatory network reconstruction from large burst sets of biological data require a proper integration technique. In this paper, we are employing Recurrent Neural Network (RNN) for modeling the dynamical behavior of gene regulatory systems and then we are applying Bayesian Model for training the RNN Model. This approach is tested against experimental data of normal and tumour prostate samples. The objective is to find regulatory interaction among genes and represent in terms of graph where genes are going to act as nodes in the graph and regulatory interactions are represented in terms directed edges.
Keywords - Bayesian Averaging Model, Gene regulatory networks, Recurrent Neural Network, regulatory interaction**.**

I. INTRODUCTION

The Genome wide gene expression microarray data sets represent a snapshot of average transcript levels for cells grown under specified conditions. A Gene Regulatory Network (GRN) is a network representing nodes as genes and causal interactions between genes as edges [4]. GRN form dynamic and distributed systems which control the expressions of the various genes in the cell. It plays a key role for solving various biological problems; for example, for identifying transcription factors of specific disease marker genes and also it can be used to investigate evolutional clues in developmental processes which affect the development of certain diseases [2]. Recently, a modern experimental technology is developing rapidly, and the required data availability is so vast. So, the accumulation of an unprecedented amount of this biological data has created the pressing need for the development of computational methods to analyze and interpret such data. In this work, we are focusing on the problem of reverse engineering the topology of gene regulatory networks (GRN) from temporal gene expression data that capture the network's dynamical behavior. Thus, various mathematical and statistical methods have been introduced to identify gene regulatory networks that can be used to infer causalities among genes.

The topology of a gene regulatory network captures a complex web of causal relationships, where connections represent regulatory interactions between genes. Here, causality refers to the regulation of the gene expression process. Gene regulatory interactions are indirect in the sense that causal influences are exerted not by the genes themselves, but rather by specialized protein products (transcription factors). Transcription factors bind to specific regions in the sequence of target genes in order to modify their expression by inducing changes in the rate of production of the corresponding proteins. Depending on the nature of these changes, the regulatory influence of transcription factors (regulators) can be classified either as activation when the rate of protein synthesis increases or as repression when the rate of protein synthesis decreases [1]. Discovering the network structure is called the problem of the reconstruction of the interactions graph. Identifying all functional parameters knowing the network structure corresponds to the traditional problem of identifying a dynamic system with a prediction aim [3].

Generally, there are two main challenges for genome-wide scale GRNs. Microarray data used in GRN inference consists of at most a few hundred samples (n) while the number of genes (g) could be up to the tens of thousands. Microarray data are characterized as massive, heterogeneous, high-dimensional, and net character in nature. While the small number of samples leads to a temporal aggregation bias, the measurement errors cause data to be noisy. Since there are difficulties in accurately estimating a gene expression value in the presence of other genes, discretized data have been used for the sake of simplicity [4]. However, in reality, gene expression levels tend to be continuous rather than discrete. Multivariate methods such as multiple regressions identified correlation between gene expression levels but were unable to determine whether the genes interact directly or indirectly through other genes thus limiting the effectiveness of GRN reconstruction.

In the proposed system, we use a Bayesian Model in order to reverse engineer the topology of a gene regulatory network from temporal data that capture the network's dynamical behavior. The recurrent neural network (RNN) formalism is employed for modeling the network's dynamics.

II. METHODS FOR GENE REGULATORY NETWORK

A. Recurrent Neural Network

The dynamical behavior of gene networks can be represented by additive regulation models, where the combined causal influence of a set of regulators to a target is expressed as the weighted sum of their expression levels. In this case, the strength and nature of the causal interaction between genes i (target) and j (regulator) are expressed as a real number wij. A positive value of wij denotes an activatory relationship (gene j activates gene i), a negative value denotes a repressive relationship (gene j represses gene i), and a zero value signifies the absence of a causal relationship between the two genes. This way, the structure of a gene network can be described by a weight matrix $W[ij]_{N \times N'}$, where N is the number of genes in the network.

A frequently used model for describing the dynamics of such a system is the recurrent neural network, whereby the expression level xi of the ith gene varies temporally according to,

$$x_i(t+\Delta t) = \frac{\Delta t}{c_i} f\left(\sum_{j=1}^{N} w_{ij} x_j(t) + b_i\right) + \left(1 - \frac{\Delta t}{c_i}\right) x_i(t)$$

where bi is a bias term that can be interpreted as the basal expression level of the ith gene and ci is a time constant that acts as a scaling factor [1]. Function f is a sigmoidal function such as the logistic function

B. Bayesian Network (BN)

A Bayesian Network is widely used as a gene network model and many computational methods have been proposed to estimate Bayesian networks from gene expression data. A major feature of a Bayesian network is that it can capture causal relationship as directed edges between genes.

BNs are especially suitable for learning regulatory networks or biological pathways for the following reasons: (1) the sound probabilistic semantics allows BNs to deal with the noises that are inherent in experimental measurements; (2) BNs can handle missing data and permit the incomplete knowledge about the biological system; and (3) BNs are capable of integrating the prior biological knowledge into the system interrelationship and then show the corresponding source of that concept [6].

C. Bayesian Model Averaging

Suppose a standardized gene expression dataset X= ($x_{1, \ldots, x_i}$) where x_i is length n vector expression for the ith gene and M_{il} is the lth regression model where the ith gene is dependent variable with a set of independent variables(in-degrees), K .

$X_{i=}$ (1)

where is an error term and is a coefficient representing the effect of gene j on the ith gene. Let be the true coefficient, then by Bayes' rule and the law of total probability,

$$p(\theta_{ji}|X) = \sum_{l=1}^{L} p(\theta_{ji}|X, M_{il}) \quad (2)$$

Where

$$p(M_{ji}|X) = p(X|M_{il})p(M_{il}) / \sum_{h=1}^{L} p(X|M_{ih})p(M \quad (3)$$

where L is the number of all possible models which include the jth gene as one of their predictor variables. These equations mean that the full posterior distribution of is a weighted average of its posterior distributions, . The weight is the posterior model probability, in (3).

III. PROPOSED SYSTEM

The proposed architecture of Reconstruction of Gene Regulatory Network is shown in figure: 1. For reconstruction of Gene Regulatory Network input is the Gene Expression Dataset. We process this data using number of steps and generate the reconstructed gene regulatory network with minimum possible errors. The reconstruction of gene regulatory network using gene expression dataset follows the following steps:

- Select the gene expression dataset.
- Indentify causal relationship and generate the temporary network from the adjacency matrix generated from dataset.
- Train the dataset using recurrent neural network which removes the noise of dataset.
- Apply the Bayesian network model to find the probability of the next value with minimum possible error.
- Again train this model using Adaptive Neuro-Fuzzy Inference System.
- Finally, construct a modified network processed through above steps.

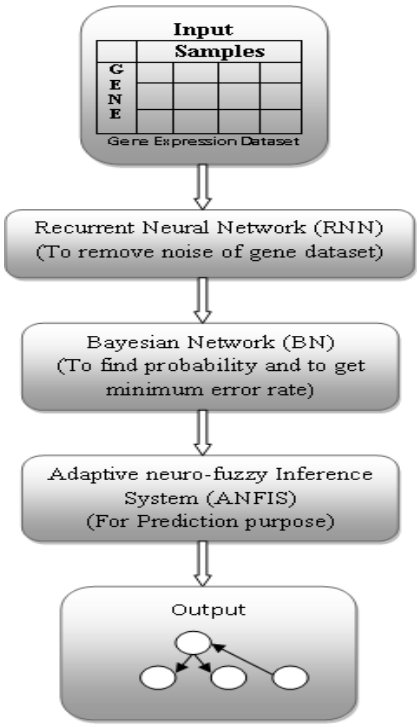

Figure 1: Architecture of Proposed System

II. IMPLEMENTATION DETAILS

The proposed system is implemented with the proposed techniques as Recurrent Neural Network and Bayesian Model. The implementation details of our proposed system are described in short in the next sections:

A. Finding Adjacency Matrix from Gene Expression Dataset

As we are taking gene expression dataset as input to the proposed system and this shown in figure 2. It contains sample, id and their value at particular interval called as time-series. As shown in figure 3, we are finding adjacency matrix which contain the value of dataset in the form of 0 and 1. This shows that, if the value is 1 then gene is activating to other and if the value is 0 then it is repressing the other.

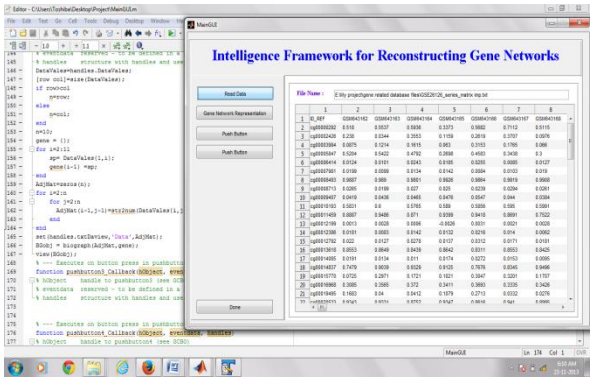

Figure 2: Gene Expression Dataset on framework

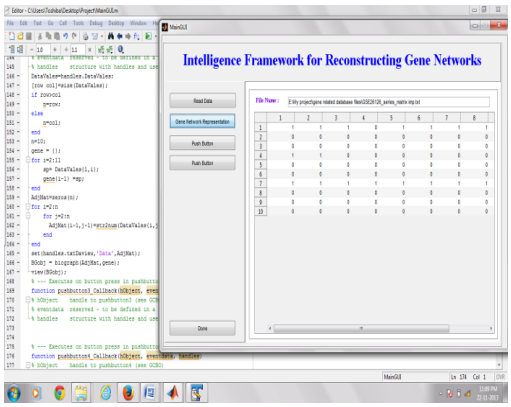

Figure 4: Representation of adjacency matrix.

From this adjacency matrix, we are generating network graph to show how one gene is activating to other one. In that, node represent gene and transition represent regulatory interaction among them. This is shown in figure 5.

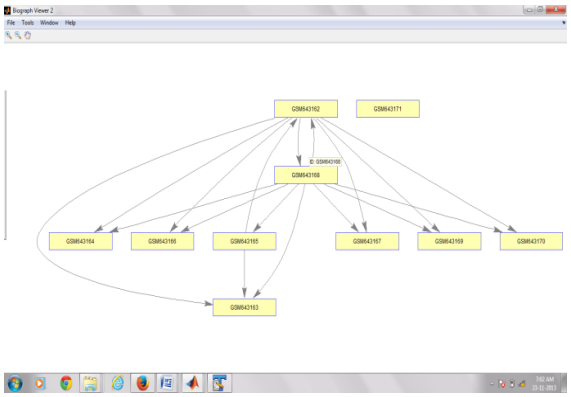

Figure 5: Representation of relationship among genes

B. Apply Recurrent Neural network

By applying a recurrent neural network, we are training the gene data set with number of parameters as shown in figure 6.

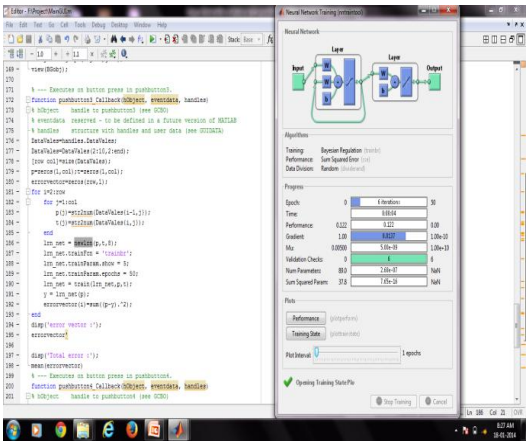

Figure 6: Neural Network Training

The training of gene expression dataset is shown in the form of graph of performance and training state. The parameters which we have taken in figure 6, on the basis of that only we are showing the graph in figure 7 and 8 respectively.

Figure 6: Experimental results of performance.

In the above figure, the results are shown for train, validation and test. The straight line for each after some epoch indicates that the output is coming same for each epoch. Figure 7 shows the individual performance of each parameter.

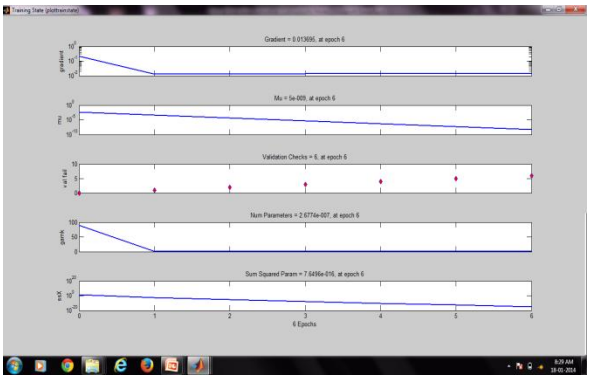

Figure 7: Experimental results of training state.

Finally, the output of recurrent neural network generate an error vector of selected gene dataset which is shown on figure 8. An error vector is calculated throgh each row and finally the total error is shown as addition of error vector.

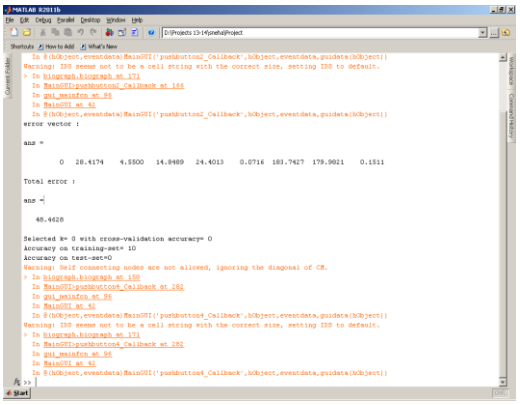

Figure 8: Error Vector generated after running RNN

C. For reconstructing the gene network apply Bayesian Model

Bayesian model finds the probability of different genes, that how will they interact with each other. This procedure is an iterative through which we find minimum possible error among genes. From figure 5 and figure 9, we can easily find the difference between two.

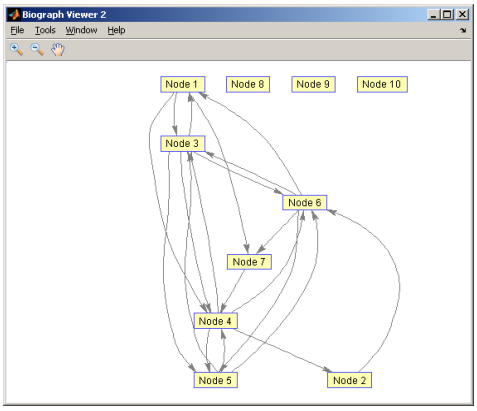

Figure 9: Reconstructed Network

III. CONCLUSION

The proposed approach for reconstructing gene regulatory network find error rate of given dataset and then calculate probability of new value with minimum possible errors. Using this concept, we generate the reconstructed graph of the used dataset. Currently, this system only works on prediction of value. The future scope is to optimize the error rate at minimum possible level.

References

[1] Kyriakos Kentzoglanakis, Matthew Poole, "A Swarm Intelligence Framework for Reconstructing Gene Networks: Searching for Biologically Plausible Architectures", Vol. 9, No. 2, March/April 2012, IEEE.

[2] Haseong Kim, Erol Gelenbe, "Reconstruction of Large-Scale Gene Regulatory Networks Using Bayesian Model Averaging", VOL. 11, NO. 3, September 2012, IEEE.

[3] Bruno-Edouard Perrin, Liva Ralaivola, Aurelien Mazurie, Samuele Bottani, Jacques Mallet, Florence d'Alche–Buc, "Gene networks inference using dynamic Bayesian networks", Bioinformatics Vol. 19 Suppl. 2 2003.

[4] Ramesh Ram and Madhu Chetty, " A Markov-Blanket-Based Model for Gene Regulatory Nework Inference", VOL. 8, NO. 2, MARCH/APRIL 2011, IEEE/ACM transactions on computational biology and bioinformatics.

[5] Yoshinori Tamada, Seiya Imoto, Hiromitsu Araki, Masao Nagasaki, Cristin Print, D. Stephen Charnock-Jones, and Satoru Miyano,"Estimating Genome-Wide Gene Networks Using Nonparametric Bayesian Network Models on Massively Parallel Computers", IEEE/ACM Transactions On Computational Biology And Bioinformatics, VOL. 8, NO. 3, MAY/JUNE 2011.

[6] Xiaotong Lin, Xue-wen Chen,"Gene Network Learning Using Regulated Dynamic Bayesian Network Methods", 2008 Seventh International Conference on Machine Learning and Applications.

[7] Derrick T. Mirikitani, Incheon Park, Mohammed Daoudi, "Dynamic Reconstruction from Noise Contaminated Data with Sparse Bayesian Recurrent Neural Networks", IEEE Computer Society-2007.

[8] M. Zou and S. Conzen, "A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data," Bioinformatics, vol. 21, no. 1, p. 71, 2005.

[9] Wei-Po Lee, Wen-Shyong Tzou, "Computational methods for discovering gene networks from expression data", Briefings In Bioinformatics. Vol 10. No 4. 408 ^ 423¸ 2009.

[10] www.ncbi.nlm.nih.gov/**geo**/

[11] Maraziotis, A. Dragomir, A. Bezerianos, "Gene Networks Inference From Expression Data Using a Recurrent Neuro-Fuzzy Approach", 2005 IEEE.

[12] Long Cheng, Zeng-Guang Hou, Yingzi Lin, Min Tan, Wenjun Chris Zhang, Fang-Xiang Wu, "Recurrent Neural Network for Non-Smooth Convex Optimization Problems with Application to the Identification of Genetic Regulatory Networks", VOL. 22, NO. 5, MAY 2011,IEEE.