International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:9, 2007

A Dynamic Time-Lagged Correlation based Method to Learn Multi-Time Delay Gene Networks

Ankit Agrawal, and Ankush Mittal

***Abstract*—**A gene network gives the knowledge of the regulatory relationships among the genes. Each gene has its activators and inhibitors that regulate its expression positively and negatively respectively. Genes themselves are believed to act as activators and inhibitors of other genes. They can even activate one set of genes and inhibit another set. Identifying gene networks is one of the most crucial and challenging problems in Bioinformatics. Most work done so far either assumes that there is no time delay in gene regulation or there is a constant time delay. We here propose a Dynamic Time-Lagged Correlation Based Method (DTCBM) to learn the gene networks, which uses time-lagged correlation to find the potential gene interactions, and then uses a post-processing stage to remove false gene interactions to common parents, and finally uses dynamic correlation thresholds for each gene to construct the gene network. DTCBM finds correlation between gene expression signals shifted in time, and therefore takes into consideration the multi time delay relationships among the genes. The implementation of our method is done in MATLAB and experimental results on Saccharomyces cerevisiae gene expression data and comparison with other methods indicate that it has a better performance.

***Keywords*—**Activators, correlation, dynamic time-lagged correlation based method, inhibitors, multi-time delay gene network.

I. Introduction

ONE of the most important objectives in the post genomic era is to learn the inter-relationships amongst genes [1], and therefore, learning gene networks has become one of the most active research areas in bioinformatics. Realization of the gene network can be highly useful for applications like validating drug targets [2], discovering the higher order structures of organisms, and interpreting their behavior [3]. A gene network can be constructed by analyzing gene expression data obtained after microarray analysis [4]. Time series gene expression data gives the expression levels of the genes at successive time points. Thus, it contains rich information about the gene interactions. It is therefore desirable to extract this information from the gene expression data and use it to construct the gene network. However, due to noise and irregularity in gene expression data [5], it is difficult to learn the exact gene network as the problem is NP-complete [6]. Although the currently available datasets provide gene expression values of a large number of genes, the data in the temporal dimension is very limited, i.e., the number of successive time measurements of a gene is far too less than the number of genes. Therefore, to estimate the gene network taking into account the multi-time delay relationships among the genes is a very challenging task, and some heuristics need to be applied to obtain the gene network efficiently and with a fairly good accuracy.

Authors are with the Department of Electronics and Computer Engineering, Indian Institute of Technology, Roorkee, India (phone: 91-1332-285713; e-mails: ank47ume@iitr.ernet.in, ankumfec@iitr.ernet.in).

Various computational methods have been used to model gene networks. Reference [7] adopted a Boolean network model of the gene network. Reference [8] learnt the gene regulations by linear regression. Reference [5] used the Bayesian networks to learn the gene network. The use of Bayesian network is extended in [1] and [9] by combining non-parametric regression [1] to detect the nonlinear relationship among genes and by making use of some biological information to improve the learning performance [9]. Reference [10] used dynamic Bayesian network and non-parametric regression model to learn the gene network. In [6], a clustering technique was employed and an objective function was subsequently used to measure the degree of activation or inhibition of a gene by another gene. They obtained a gene regulatory network where activator and inhibitor clusters were found. This technique, however, considered that a gene can either be in an activator cluster or in an inhibitor cluster. This might not always be the case as a gene may activate one gene and inhibit another.

Recently, a few important facts relating to learning gene networks have been discovered. When a gene regulates another gene, there is a time delay between the changes of expression levels of the genes [3, 11]. Time delay in gene regulation results due to the delays in various related biological processes like transcription, translation, transport, etc. Researchers have tried to incorporate time delay into their models and assumed that the time delay is constant. Based on this assumption, a linear model was used to learn the gene network in [8]. References [12] and [13] used the dynamic Bayesian network (DBN) to model the time delay in the gene network. Some research [3] has shown that different gene pairs have different time delays for gene regulation. Reference [14] used Bayesian network framework and introduced a new structure learning algorithm to learn the multi-time delay gene

network. In [15], a supervised learning approach was used to learn the gene network by building decision-tree-related classifiers, which predict gene expression from the expression data of other genes. Reference [11] uses a mixed integer linear programming framework for inferring time delay in gene regulatory networks.

Already signal processing techniques have widely been used in bioinformatics. References [16] and [17] provide a good review of the use of signal processing concepts in genomics and proteomics, and genomics signal processing. In this work, the focus is on the use of signal processing techniques on gene expression data. Most of the work till now on gene expression data is limited to detecting gene clusters, and little work has been done to find the gene network. Reference [18] used signal processing metrics like power spectral density, coherence, transfer gain, and phase shift to find the similarity in time series gene expression data. In [19], the gene network was modeled using multi-criterion optimization. They imposed various constraints on the genetic network model, based on biological knowledge about real genetic networks like limited connectivity, redundancy, stability and robustness, trying to cope up with the problem of less data. But they did not address the multi-time delay relationships among genes. Reference [20] used graphical Gaussian modeling and standard multivariate statistical techniques to deduce regulatory relationships from gene expression data. But it also did not consider the multi-time delay relationships among the genes. Therefore, most of the work done so far has not taken the specific advantage of the information hidden in the temporal aspect of data that is provided by the time-series gene expression data [18].

Correlation techniques also have been used in development of enzymatic pathways, and genetic interaction networks. Reference [21] used the Correlation Metric Construction (CMC) approach to model the reaction pathway for glycolytic biochemical system. It uses a time-lagged correlation metric as a measure of distance between reacting species. It had some drawbacks; still the example showed that even when the specific method of interaction is unknown or unmeasured, useful information could be inferred about the overall structure of a network from forced dynamic experiments [22]. In [22] also, the same concept of forced dynamic experiments was used to monitor gene transcription in response to a time-varying input light intensity signal. It found correlation between the time-lagged profiles of genes, and the input light intensity signal.

In short, the field of gene networks is still not fully explored, and especially the area of multi-time delay gene networks needs directed research efforts. This work uses correlation techniques to analyze the gene expression data and solve the multi time delay gene network problem. Correlation between time-shifted expression values of different pairs of genes is calculated and the potential activator & inhibitor relationships between the genes are estimated. Post-processing of these potential relationships is also proposed to remove the false relationships between genes due to a common parent. Finally, dynamic correlation thresholds for each gene are used to determine the final relationships among genes.

The rest of the paper is organized as follows. In Section 2, a brief overview of signal processing is presented with special focus on correlation techniques. The major contribution of our paper is discussed in Section 3, which describes the multi-time delay gene network model followed by the proposed dynamic time-lagged correlation based method (DTCBM) to obtain the multi-time delay gene network. Experimental results and comparison on two real datasets of yeast are presented in Section 4, followed by the conclusion in Section 5.

II. Signal Processing Preliminaries

A. Gene Expression as a Discrete Time Signal

A discrete time signal $x[n]$ is a set of measurements x made at discrete evenly spaced time points n. The function $x[n]$ is defined only for integer values of n. The gene expression values of a gene can be treated as a discrete time signal. Therefore, we have as many discrete time signals as the number of genes. Various signal processing metrics can then be applied on these signals to find the time delayed dependencies between them, which can give the knowledge of the gene network. Here, these signals are assumed to be time-invariant, i.e., although the measurements of the expression values change over time, but their variability does not change [18]. This means that the behavior of a gene does not change with time. If it regulates some gene, it will always do so under similar conditions. This is an important assumption, given the fact that the biological systems are highly dynamic and can be infinitely complex in nature. Still, this is assumed correct, although it is difficult to be proved. Here, we do not intend to prove it, rather we will stick to the assumption that the signals of gene expression are time invariant.

B. Correlation Techniques

Correlation is used to determine the extent to which two signals are related, either positively or negatively. The result of correlation is expressed as correlation coefficients, whose value can lie between -1.0 to 1.0. A value closer to -1.0 or 1.0 indicates strong dependence between the signals, negatively and positively respectively. A value close to 0.0 suggests that the signals are independent.

Let x[n] and y[n] represent two discrete time signals. The true cross-covariance $\phi_{xy}(m)$ of these signals is the cross-correlation of the mean-removed sequences:

$$\phi_{xy}(m) = E[(x_{n+m} - \mu_x)(y_n - \mu_y)^*] \quad (1)$$

where $E[\square]$ is the expected value operator, μ_x and μ_y are the mean values of the two stationary random processes:

$$\mu_x = \frac{1}{N}\sum_{i=0}^{N-1} x_i \text{ ; and } \mu_y = \frac{1}{N}\sum_{i=0}^{N-1} y_i \quad (2)$$

Therefore, cross-covariance at a time lag of m time points is:

$$\phi_{xy}(m) = \sum_{n=0}^{N-|m|-1} \left(x_{n+m} - \frac{1}{N} \sum_{i=0}^{N-1} x_i \right) \left(y_n^* - \frac{1}{N} \sum_{i=0}^{N-1} y_i^* \right) \quad (3)$$

Finally, the correlation coefficient of sequences x[n] and y[n] with a time lag of *m* is obtained as:

$$C_{xy}(m) = \frac{\phi_{xy}(m)}{\sqrt{\phi_{xx}(0)\phi_{yy}(0)}} \quad (4)$$

The squared correlation coefficient is the proportion of variance in y[n] that can be accounted for by knowing x[n]. Conversely, it is the proportion of variance in x[n] that can be accounted for by knowing y[n].

III. Correlation Analysis of Gene Expression Data

A. The Multi-Time Delay Gene Network Model

In general, gene regulation is described as follows: gene g_1 is said to regulate gene g_2 positively if g_1 transcripts to mRNA r_1 and generates a specific protein p1, which activates the expression of g_2 and thus increases the expression of g_2. Similarly, g_1 regulates g_2 negatively if the expression level of g_2 reduces with increase in expression of g_1. It is well known that there exists some finite time delay between the expression of an activator/inhibitor gene and the expression of the gene which it is regulating. The time delay intervals between their expression may also be different for different gene pairs. It is further known that there should be a maximum time delay interval in a gene network since the time of a cell cycle is limited [14]. The gene expression datasets that are available presently capture the gene expression values of genes every 10-30 minutes. In general, an activator/inhibitor gene can regulate another gene either instantly, or in the next time slice or after two time slices or up to $\tau_{\max}$ time slices. The parameter $\tau_{\max}$ indicates the maximum delay within which a gene can regulate another gene. So the gene expression signal of each gene is correlated with other expression signals of other genes with a maximum lag of $\tau_{\max}$ time slices.

An important advantage of this gene network model is that it does not require any discretization thresholds. It is well known that discretization leads to loss of information. This is one of the major defects of the Boolean model of gene network, where the gene can be either active or inactive. By discretizing, significant change in expression value can be missed. Also, an increase in expression value of a gene in the inactive state can be misinterpreted, if the gene is still inactive according to the discretization thresholds, and vice-versa. Therefore, better approach is to focus on the relative change in the expression values, rather than absolute values.

This correlation based approach places the limitation that any gene cannot regulate itself. This is because the correlation coefficient of each gene expression signal with itself, at zero time lag, i.e., $C_{ii}(0)$ for gene g_i, will always be 1.00, implying that each gene is an activator of itself. Therefore, such interactions are not considered and it is assumed that genes are not self regulated. This is reasonable also since if a gene was to regulate itself, either because of positive feedback its expression value will go on increasing indefinitely (in the case of activation), or because of negative feedback, its expression value will remain constant (in case of inhibition). The first case is certainly impractical, and the genes belonging to the second case do not appear to be involved in regulation, as their expression levels do not change significantly. Thus, they are not of interest for estimation of the gene network, and will be filtered out initially.

To summarize, the model takes into consideration the following features of the gene network:

1) Genes can have more than one activators and inhibitors.
2) No gene can be an activator or inhibitor of itself.
3) Maximum delay in gene regulation can be $\tau_{\max}$ time slices.
4) A gene can activate one gene and inhibit another

B. Dynamic Time-Lagged Correlation based Method

DTCBM has five stages viz., preprocessing, finding the time-lagged correlation matrix, applying decision rule to get potential activators and inhibitors, post-processing, and applying dynamic correlation thresholds to get final activators and inhibitors.

1) *Preprocessing:* The genes that have too many missing values in the dataset are removed. And, for a single missing value between two time slices, linear interpolation is used. Then, the genes that show very little variation over time are filtered away as it indicates that they are not involved in the process of regulation. For filtering, the criteria used is that the standard deviation of the expression level of the gene over all time slices must be greater than dev_{min}, i.e.,

$$dev = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \geq dev_{min} \quad (5)$$

$$dev_{min} = 0.01 \text{ x } range \quad (6)$$

where
x_i is the expression value of the gene at i^{th} time slice;
$\bar{x}$ is the average value of the gene over all time slices;
n is the number of time slices; and
range is the range of expression values in the dataset.

After preprocessing step, we have the gene expression matrix, $E[m \times n]$, where m is the number of genes and n is the number of time slices.

2) *Finding time-lagged correlation matrix:* Correlation between all ordered pairs of genes with all allowed time lags is calculated and stored in a 3-dimensional array, $C[m \times m \times (\tau_{\max} + 1)]$. The entries of matrix C are obtained as follows:

$$C[i, j, k] = C_{ij}(-k) \quad (7)$$

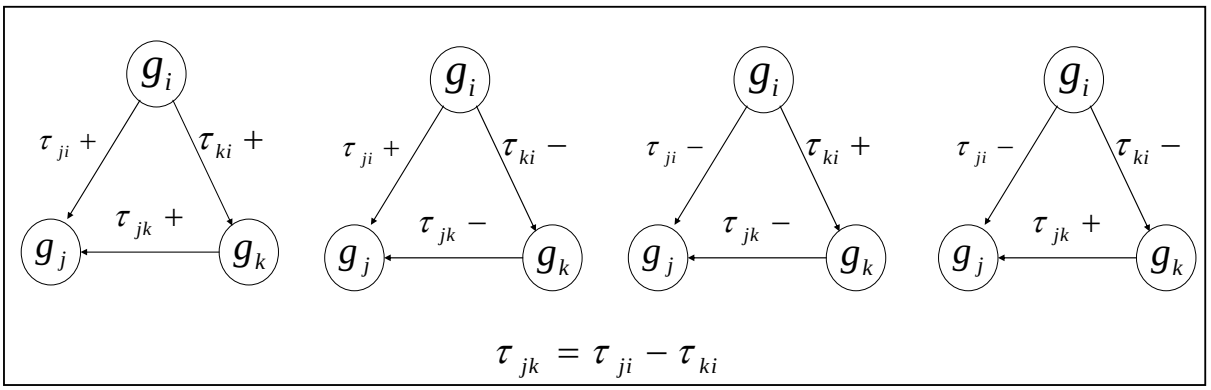


Fig. 1 False edge between g_j and g_k because of a common parent g_i. The label on the edges represents the time delay, and the sign indicates whether it is an activator (+) or inhibitor (-)

$$\forall\ 1 \le i \le m,\ \ 1 \le j \le m,\ 0 \le k \le \tau_{\max}$$

where $C_{ij}(-k)$ is defined as in (4).

The entry $C[i, j, k]$ is the correlation coefficient of the expression signal of gene g_i shifted left by k time slices and, the expression signal of gene g_j. If this coefficient is high (may be positive or negative), it implies that gene g_j regulates gene g_i.

3) *Applying decision rule to get potential activators and inhibitors:* After the time-lagged correlation matrix C is obtained, it is analyzed for the activators and inhibitors of each gene. The 2-dimensional matrix $C[i,:,:]$ contains all the correlation information of the left time shifted signal of gene g_i with all other genes. Therefore, it has all the information about the regulators of gene g_i. The following decision rule is used to find the potential regulators of gene g_i: gene g_j is considered to be a potential regulator of gene g_i with a time delay of k time slices, if the maximum(minimum) of the one-dimensional array $C[i, j,:]$ is $C[i, j, k]$, and it is greater(smaller) than a threshold correlation $\lambda_{\min}(-\lambda_{\min})$, i.e.,

gene g_j activates gene g_i after k time slices if,

$$C[i, j, k] = \max(C[i, j,:]) \ge \lambda_{\min} \quad 0 \le k \le \tau_{\max} \qquad (8)$$

gene g_j inhibits gene g_i after k time slices if,

$$C[i, j, k] = \min(C[i, j,:]) \le -\lambda_{\min}\ 0 \le k \le \tau_{\max} \qquad (9)$$

where $\lambda_{\min}$ is a predefined threshold discussed later in this section.

4) *Post-processing:* The relationships obtained by the previous stage may contain many false positives. Two genes will appear to have high correlation if both have same parents. E.g. if gene g_i regulates both g_j and g_k with a time delay of τ_{ji} and τ_{ki} respectively, then g_j and g_k will have high correlation with a time delay equal to the difference of τ_{ji} and τ_{ki}. So, g_k might be considered an regulator of g_j. If the time delay and nature (activation/inhibition) of this relationship is in accordance with the above two relationships and its correlation coefficient is less than $\lambda_{\max}$, it is removed in this stage. Fig. 1 shows three genes having such a relationship. Mathematically,

Find all gene triplets $g_i, g_j,$ and g_k, such that

$$\tau_{jk} = \tau_{ji} - \tau_{ki} \text{ and } \quad v_{jk} = v_{ji} \times v_{ki} \qquad (10)$$

$$\text{if } \max(|C[j, k,:]|) < \lambda_{\max}$$

then remove the edge $g_k \to g_j$

where $v_{ij} = \begin{cases} +1 & \text{if } g_j \text{ activates } g_i \\ -1 & \text{if } g_j \text{ inhibits } g_i \end{cases}$

$\lambda_{\max}$ is a predefined threshold discussed in the next subsection.

5) *Dynamic correlation thresholding:* There are several methods of determining correlation thresholds:

- *Pre-selected thresholds*: The thresholds can be made fixed irrespective of the data used, as used in [22].
- *Data dependent thresholds*: Thresholds depend on the dataset. For example, mean of the data values.

Both these methods have their limitations. While the first method is totally insensitive to the type if data used, the second one may become too sensitive to the data, and both can lead to inaccurate results. Used here is a hybrid of the two approaches, where the thresholds are dependent on the data within a range. We use two fixed thresholds $\lambda_{\min}$ and $\lambda_{\max}$, with $\lambda_{\min} < \lambda_{\max}$. For determining the

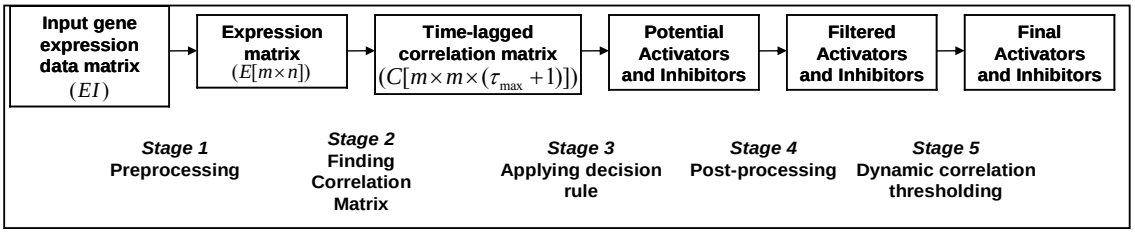


Fig. 2 Block Schematic of the Dynamic Time-Lagged Correlation based Method

regulators of genes, a correlation threshold λ_i is determined for each gene, which depends on its correlation coefficients with its potential parents up to stage 4. λ_i is the mean of these correlation coefficients. This final threshold is intended to be kept between $\lambda_{\min}$ and $\lambda_{\max}$, and therefore if the calculated threshold exceeds $\lambda_{\max}$, it is rejected and $\lambda_{\max}$ is used for the decision rule. Mathematically,

$$\lambda = mean(C')$$

where $C' = \{\max(|C[i,j,:]|) \mid \text{edge } g_j \rightarrow g_i$ retained after post-processing$\}$

$$\lambda_i = \begin{cases} \lambda & \lambda \le \lambda_{\max} \\ \lambda_{\max} & \lambda > \lambda_{\max} \end{cases} \quad (11)$$

The block schematic of various stages in the DTCBM is presented in Fig. 2. The implementation of the DTCBM is done in MATLAB, and its time complexity is $O(m^3 + m^2 n)$, where m is the number of genes, and n is the number of time slices in the dataset.

IV. Experiments and Results

To evaluate the performance of the proposed method, the *Saccharomyces cerevisiae* cell cycle gene expression data [23] was analyzed. Here, the *cdc15* dataset was used as it has the maximum number of gene expression measurements with constant time interval between them (10 minutes). We used two set of genes and estimated the subnetwork for them, and compared them with networks already learnt so far. For all experiments, the parameters were set as following:

The maximum time delay, $\tau_{\max}$ is set to 4, i.e., 40 minutes, since the time of single cell cycle of *S. Cerevisiae* is about 1.5 hours [14]. The lower correlation threshold $\lambda_{\min}$ is set to 0.70, and the upper correlation threshold $\lambda_{\max}$ is set to 0.80.

For comparison of results, the following parameters are used:

Accuracy, $\alpha = \frac{\omega}{\xi}$, which is a measure of the capability of the algorithm to recover the correct edges.

Precision, $\rho = \frac{\omega}{\omega + \chi}$, which is the chance that a edge detected by an algorithm is correct.

where ω indicates the number of true positives;
χ indicates the number of false positives; and
ξ indicates the total number of edges in the target network

A. Experiment 1

First, the algorithm was tested with the set of genes used in [10] to find the subnetwork and compared it with the actual network and the one published in [10] using DBNNR method. It included the following fourteen genes: FUS3, FAR1, SWI4, SWI6, CLN1, CLN2, CDC28, CLN3, MBP1, SIC1, CLB5, CLB6, CDC20, and CDC6. The target subnetwork (registered in KEGG [24] database), the one obtained in [10], and the one estimated by DTCBM, are shown in Fig. 3. The edges in the dotted circles can be considered as correct as they represent co-regulated clusters. The number of edges in the target network is not known exactly, as edges from one cluster to another implies that edges from all genes of first cluster to all genes of second one are also correct. Therefore, accuracy is difficult to be determined. Therefore, the number of correct edges recovered is taken as a measure of accuracy. The comparison of DTCBM with DBNNR indicates that DTCBM was able to capture more true positives, thereby having a higher accuracy, and simultaneously, the precision was also better. The results in tabular form are presented in Table I.

B. Experiment 2

As a second experiment with another set of genes, the comparison of the subnetwork obtained was performed with the one published in [15]. It included the following genes: CLN2, CLN1, CDC20, MBP1, SWI5, CLB2, SKP1, CLN3, CLB1, CLB6, CLB5, and CDC34. The network published in [15], and the one estimated by DTCBM are shown in Fig. 4. DTCBM was able to capture 26 edges, out which 19 were correct edges. There were a total of 30 edges in the target network. Not only most of the edges were detected accurately, the edge-characteristics, i.e., activation/inhibition and time delay also matched with that found in [15]. Overall, the results indicate that DTCBM scores an accuracy of 63% and precision of 73% on this dataset, as shown in Table II.

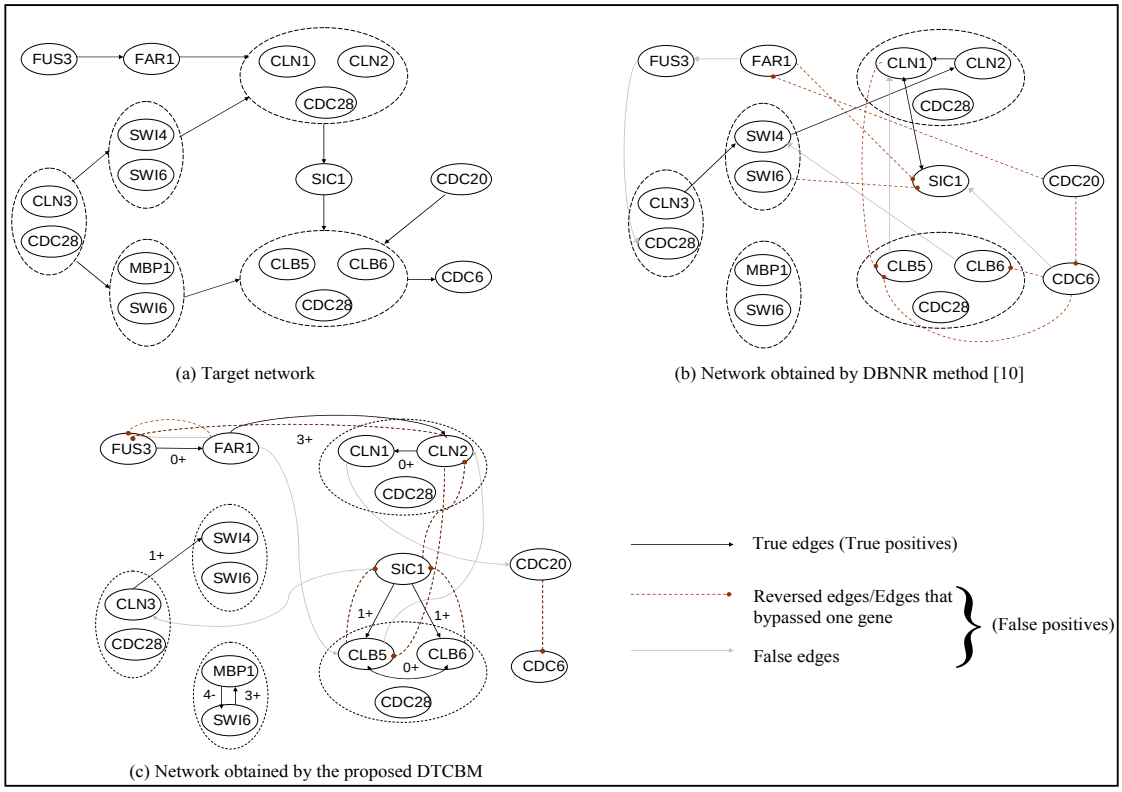


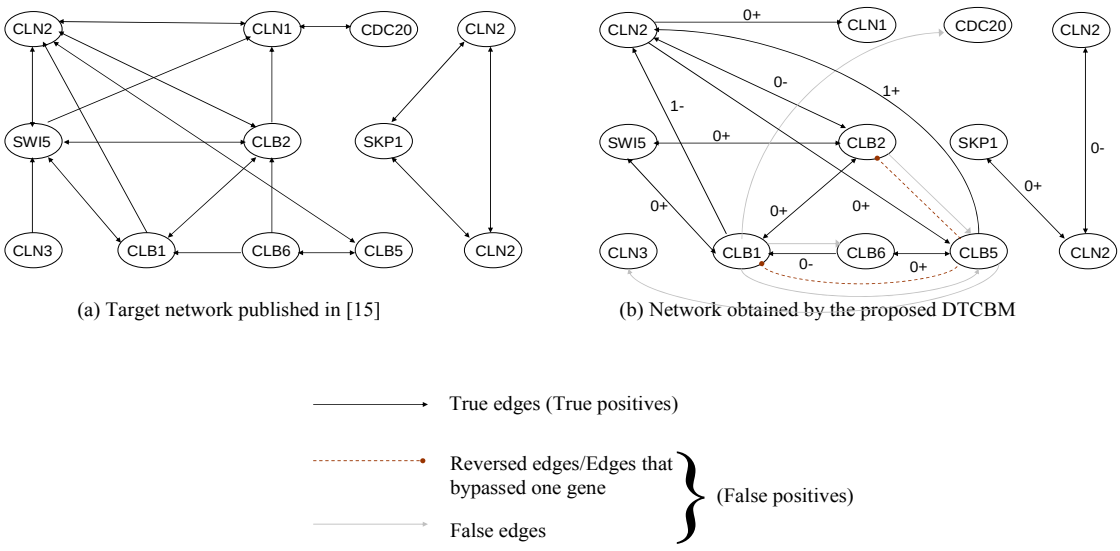
Fig. 3 Gene networks obtained during experiment 1. (a) The real sub network registered in KEGG [24]. (b) The gene network structure obtained by DBNNR [10] (c) The gene network structure obtained by DTCBM. Black arrows represent true edges and grey arrows represent false edges. Brown edges with a rounded head represent reversed edges or edges that bypassed one gene. The number on a true edge represents the time delay and the sign indicates whether it is an activator (+) or inhibitor (-). A double headed arrow is equivalent to two opposite single headed arrows)

TABLE I
EXPERIMENT 1 RESULTS

Algorithm	**True positives(ω)**	**False**	**Positives(χ)**	**Precision (ρ)**
		Reversed edges/one gene bypassed	False edges	
DBNNR	5	7	5	0.294
DTCBM	10	7	4	0.476

TABLE II
EXPERIMENT 2 RESULTS

Algorithm	**True positives (ω)**	**False**	**Positives(χ)**	**Accuracy (α)**	**Precision (ρ)**
		Reversed edges/one gene bypassed	False edges		
DTCBM	19	2	5	0.633	0.730



(a) Target network published in [15] (b) Network obtained by the proposed DTCBM

Fig. 4 Gene networks obtained during experiment 2. (a) The target sub network published in [15] (b) The gene network structure obtained by DTCBM. Black arrows represent true edges and grey edges represent false edges. Brown edges with a rounded head represent reversed edges or edges that bypassed one gene. The number on a true edge represents the time delay and the sign indicates whether it is an activator (+) or inhibitor (-). A double headed arrow is equivalent to two opposite single headed arrows)

V. Conclusion

The dynamic time-lagged correlation based method proposed in this paper estimates the multi-time delay gene network using correlation between gene expression signals shifted in time. Other contributions of the paper are a post-processing stage to remove false gene interactions and a dynamic thresholding system which allows the correlation threshold to vary in a fixed range. One of the main advantages of the algorithm is that the gene expression data need not be discretized, which eliminates the need of any pre-defined discretization thresholds. Experimental results and comparison with other gene network learning algorithms indicate that the DTCBM has a better performance both in terms of accuracy and precision.

However, the present work has some limitations as well. Here, linear correlation is used assuming that the gene regulatory relationships are linear, which may not be true. Also, the high correlation is interpreted as causation, which may not be necessarily true always. Another point is that correlation does not indicate the direction of causation. But one of the most important conditions required to interpret high correlation as causation is met in this work, which is that the caused event must take place after the causing event. This is achieved because the algorithm finds the correlation between the time-shifted gene expression signals. This also takes care of the direction of the causation. Still, the presence of false positives with high correlation indicates that the method is not perfect, although many false positives are removed in the post-processing stage. Therefore, additional inputs may be required, either in the form of more expression data, or some biological information. We would like to work in this direction in future. Future work also includes using non-linear correlation for the problem, and designing a more robust correlation threshold determining system. Combining datasets of unequal time slice spacing is also an important issue.

Acknowledgment

The first author would like to thank Mr. Ravi Gupta, IIT Roorkee for his insightful suggestions and support for this work.

References

[1] S. Imoto, T. Goto, and S. Miyano, "Estimation of Genetic Networks and Functional Structures Between Genes by Using Bayesian Networks and Nonparametric Regression," *PSB*, volume 7, pp. 175–186, 2002.

[2] C. J. Savoie, S. Aburatani, S. Watanabe, Y. Eguchi, S. Muta, S. Imoto, S. Miyano, S. Kuhara, and K. Tashiro1. Use of Gene Networks from Full Genome Microarray Libraries to Identify Functionally Relevant Drug-affected Genes and Gene Regulation Cascades. *DNA Research,* Volume 10, pp. 19-25, 2003.

[3] T. Chen, H.L. He, and G.M. Church. Modeling Gene Expression with Differential Equations. In *PSB*, vol. 4, pp. 29–40, 1999a.

[4] L.F.A. Wessels, E.P.V. Someren, and M.J.T. Reinders. A Comparison of Genetic Network Models. In *PSB*, volume 6, pp. 508–519, 2001.

[5] N. Friedman, M. Linial, I. Nachman, and D. Peer. Using Bayesian Networks to Analyze Expression Data. In *RECOMB*, pp. 127–135, 2000.

[6] T. Chen, V. Filkov, and S.S. Skiena. Identifying Gene Regulatory Networks from Exprimental Data. In *RECOMB*, pp. 94–103, 1999b.

[7] T. Akutsu, S. Miyano, and S. Kuhara. Identi.cation of Genetic Getworks from a Small Number of Gene Expression Patterns Under the Boolean Network Model. In *PSB*, pages 17–28, 1999.

[8] E.P.V. Someren, L.F.A. Wessels, and M.J.T. Reinders. Linear Modeling of Genetic Networks from Experimental Data. *ISMB*, pp. 355–366, 2000.

[9] S. Imoto, T. Higuchi, T. Goto, K. Tashiro, S. Kuhara, and S. Miyano. Combining Microarrays and Biological Knowledge for Estimating Gene Networks via Bayesian Network. In *Proceedings of 2nd Computational Systems Bioinformatics, CSB,* pp. 104-113, 2003.

[10] SunYong Kim, Seiya Imoto, and Satoru Miyano, Dynamic Bayesian Network and Nonparametric Regression for Nonlinear Modeling of

Gene Networks from Time Series Gene Expression Data . Biosystems 2004, 75, pp. 57-65, Jul. 2004.
[11] M. S. Dasika, A. Gupta and C. D. Maranas, A Mixed Integer Linear Programming (MILP) Framework for Infering Time Delay In Gene Regulatory Networks. *Pac. Sym. Biocomput.* pp. 474-485, 2004.
[12] K. Murphy and S. Mian. Modelling Gene Expression Data Using Dynamic Bayesian Networks. *Technical Report*, Computer Science Division, University of Berkekey, C.A. 1999.
[13] L. Gransson and T. Koski. Using a Dynamic Bayesian Network to Learn Genetic Interactions. *Technical Report*, 2002.
[14] Tie-Fei Liu, Wing-Kin Sung, Ankush Mittal. Learning Multi-Time Delay Gene Network Using Bayesian Network Framework. In Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2004), 2004.
[15] Lev A Soinov, Maria A Krestyaninova and Alvis Brazma, Towards reconstruction of gene networks from expression data by supervised learning. *Genome Biology* 2003, 4:R6, 2003.
[16] P. P. Vaidyanathan and Byung-Jun Yoon, The role of signal-processing concepts in genomics and proteomics. Invited paper, Journal of the Franklin Institute, special issue on Genomics, 2004.
[17] Astola, Jaakko, Edward Dougherty, Ilya Shmulevich, and Ioan Tabus, Genomic signal processing, *Signal Processing*, volume 83, number 4 pp. 691-694, 2003.
[18] A.J. Butte, Ling Bao, Ben Y. Reis, Timothy W. Watkins, and Issac S. Kohane, Comparing the Similarity id Time-Series Gene Expression Using Signal Processing Metrics. Journal of Biomedical Informatics, 34, pp. 396-405, 2001.
[19] Someren, E.P. van, L.F.A. Wessels, E. Backer, and M.J.T. Reinders, Multi-criterion optimization for genetic network modeling, *Signal Processing*, volume 83, Issue 4 pp. 763-775, 2003.
[20] Aburatani, Sachiyo, Satoru Kuhara, Hiroyuki Toh, and Katsuhisa Horimoto, Deduction of a gene regulatory relationship framework from gene expression data by the application of graphical Gaussian modeling, , *Signal Processing*, volume 83, number 4, pp. 777-788.
[21] 1. A. Arkin, P. D. Shen, and J. Ross, .A Test Case of Correlation Metric Construction of a Reaction Pathway from Measurements,. Science 277, 1275.1279 (1997).
[22] William A. Schmitt Jr., R. Michael Raab, and Gregory Stephanopoulos. Elucidation of Gene Interaction Networks Through Time-Lagged Correlation Analysis of Transcriptional Data. *Genome Research*, 14:1654–1663, 2004.
[23] P.T. Spellman, G. Sherlock, and B. Futcher. Comprehensive Identification of Cell Cycle-Regulated Genes of the Yeast Saccharomyces Cerevisiae by Microarray Hybridization. *Molecular Biology of the Cell*, 9, pp.3273–3297, 1998.
[24] http://www.genome.ad.jp/kegg/

Ankit Agrawal is pursuing his B. Tech. degree in Computer Science and Enggineering, Indian Institute of Technology. At present, he is topper of final year batch consisting of more than 350 students. His interests include Bioinformatics, data mining and Signal processing.

Ankush Mittal received the B. Tech and Masters (by Research) degrees in computer science and engineering from the Indian Institute of Technology, Delhi. He received the PhD degree from the National University of Singapore in 2001. Since October 2003, he has been working as assistant Professor in Indian Institute of Technology -Roorkee. Prior to this, he was serving as a faculty member in the Department of Computer Science, National University of Singapore. His research interests are in multimedia indexing, machine learning, Bioinformatics, E-Learning and motion analysis.