



Research Article

## A multi-objective evolutionary approach to reconstruct gene regulatory network using recurrent neural network model

Sumon Ahmed<sup>1\*</sup>, Md. Nurul Ahad Tawhid<sup>1</sup>, Kazi Sakib<sup>1</sup>, Md. Mustafizur Rahman<sup>2</sup>

1. *Institute of Information Technology, University of Dhaka*
2. *Department of Computer Science and Engineering, University of Dhaka*

### **\*Corresponding author**

Sumon Ahmed, *Institute of Information Technology, University of Dhaka, Dhaka – 1000; email: [sumon@du.ac.bd](mailto:sumon@du.ac.bd)*

Published: 13-07-2015

Biojournal of Science and Technology Vol.2:2015

Academic Editor: Dr. Md Saiful Islam

Received: 30-04-2015

Accepted: 14-06-2015

Article no: m140007

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### **Abstract**

With the advent of various data assaying techniques, gene expression time series data have become a useful resource to investigate the complex interactions occurring amongst the transcription factors and genes. While a number of methodologies have been developed to describe Gene Regulatory Network (GRN), the presence of high noise in gene expression data have made the estimation of non-linear interactions among the genes an ill-posed one. In this work, a multi-objective evolutionary strategy has been proposed to efficiently reconstruct the skeletal structure of the biomolecular network using the Recurrent Neural Network (RNN) formalism. Moreover, this work presents a second criterion for model evaluation to exploit the sparse and scale free nature of GRN. This evaluation criterion systematically adapts the max-min in-degrees to effectively narrow down the search space, which reduces the computation time significantly and improves the model accuracy. The two well-known performance measures applied to the experimental studies on synthetic network with expression data having different noise-levels. The experimental results clearly demonstrate the suitability of the proposed method in capturing gene interactions correctly with high precision even with noisy time-series data. The experiments carried out on analyzing well-known real expression data set of the SOS DNA repair system in *Escherichia coli* show a significant improvement in reconstructing the network of key regulatory genes.

**Keywords:** Gene Regulatory Network, Recurrent Neural Network, Multi-Objective Evolutionary Algorithm, Differential Evolution, Reverse Engineering

## INTRODUCTION

In recent years, the availability of large scale gene expression data has significantly increased the study of the relationship among genes. Gene expression data, whether in time-course format or steady state format open the door to the researchers to observe the interaction among thousands of genes simultaneously under various environmental conditions. Given that large volume of gene expression data is available, in principle it is possible to reverse engineer the detailed quantitative model of the biological network that adequately represents the dynamics of the underlying system (Noman et al., 2013).

Several common practical issues that make the reconstruction process of GRN difficult are small sample size compared to the number of genes, the presence of biological and experimental noise, lack of adequate knowledge of the complex dynamics and nonlinear nature of biological systems. In spite of many technical advances, the gene array technologies are still unable to acquire the quality and quantity of data that is required to capture the precise mechanism in common regulatory pathways (Noman and Iba, 2007, Schena M, 2013). Two major challenges faced by all inference methodologies (Differential Equations, Bayesian Network etc.) in terms of representation accuracy and computational feasibility while reconstructing GRN are 1) detecting the sparse topological architecture of biological network and 2) estimating the regulatory parameters from a limited amount of gene expression data corrupted with a significant level of noise. Generally, with the increase of problem dimension due to large number of genes in network, search complexity increases very rapidly and locating the global optimum solution becomes very difficult.

In order to apply a computational approach to reconstruct GRN from experimental time-series data, a mathematical model is necessary that will adequately formalize the process of gene regulation. The analysis of gene expression networks and metabolic pathways has resulted in

various types of GRN models which vary in terms of the details of biochemical interaction incorporated, discrete or continuous expression level used, deterministic or stochastic approach applied, and so forth (Noman and Iba, 2007). Among the mathematical models of GRN, Boolean Network (Sahoo et al., 2013), Linear Model (Dhaeseleer et al., 2013), Bayesian Network (Mazur and Kaderali, 2013), Neural Network (Vohradsky J, 2013), Differential Equations (Chen et al., 2013), Linear Time-Variant Model (Kabir et al., 2013), S-system Model (Savageau M A, 2013) and models including stochastic components on the molecular level (McAdams and Arkin, 2013) are well known. Boolean Networks and Linear Models are simplistic approaches that employ pairwise association measures such as conditional mutual information for inferring the interactions between genes (Chowdhury et al., 2013). Having low computational complexity, these methods can easily scale up to very large networks of thousands of genes (Basso et al., 2013). Bayesian networks (BN) are based on the strong foundation of probability and statistics where directed edges and conditional probability distributions are used to represent dependencies between nodes. Differential Equation (DE) is a member of sophisticated and well established class of models that maintains a balance between model complexity and mathematical tractability. Several linear and non-linear types of DE models such as Linear Time-Variant Model, S-system Model, etc. have the ability to depict system dynamics in continuous time (Chowdhury et al., 2013).

In this work, the Recurrent Neural Network (RNN) model (Wahde and Hertz, 2013) along with a natural computational method is used to extract regulatory interactions among genes from gene expression data sets. Among the reconstruction approaches applied to infer GRN, the RNN model is of particular interest because of its capability to adequately discover the nonlinear and dynamic interactions among the genes (Noman et al., 2013, Wahde and Hertz, 2013, Chiang and Chao, 2013). With the network of nonlinear processing

elements, the model can reasonably capture various dynamics and mechanisms that could be present in a complex biological system. However, inferring GRN using RNN model demands the estimation of large parameter sets that also increase with the number of genes present in the target network. Thus the method may get stuck on some locally optimum solution and fail to predict the true skeletal structure in case of larger biological networks. To overcome this problem, the proposed methodology incorporated another objective function that is calculated by summing up the number of regulatory inputs of all the genes in the system (Ahmed et al., 2013). As the most biological systems are sparse (Noman et al., 2013, Noman and Iba, 2007), the smaller values of this second objective function ensure the biological reality in inferred gene regulatory networks.

Applying a mathematical model for inferring GRN requires the development of some algorithmic techniques that will estimate the values of model parameters. Some algorithmic techniques such as particle swarm optimizations (Sultana et al., 2013), evolutionary algorithms (Noman et al., 2013, Noman and Iba, 2007), etc. have already been developed in the field of computational intelligence and machine learning that help the biologists to form new hypothesis about the biological systems (Noman et al., 2013) and to design new experiments. In this work, an *Evolutionary Algorithm (EA)* based inference technique using Recurrent Neural Network (RNN) model has been used with the aim of providing a method that can fulfill the experimental requirements.

As the proposed methodology uses more than one fitness function, a natural multi-objective computational approach known as elitist *Differential Evolution for Multi-Objective Optimization (DEMO)* is used. DEMO, belonging to the group of evolutionary algorithms, is proven to be very effective in solving different conflicting multi-objective optimization problems arising in different domains (Ahmed et al., 2013). Among

the family of multi-objective evolutionary algorithms, the proposed methodology has the unique feature of self adaptation. Based on its objective functions, the algorithm converges rapidly without the need of setting any threshold values on the interactions of a particular gene.

The inference capability of the proposed method has been highlighted in different learning experiments using both artificial and real gene network data. Artificial network data with varying noise levels and characteristics were chosen and simulated to obtain synthetic time-series data set and the underlying skeletal network architecture. The reconstruction results depict the suitability of the proposed approach as it correctly identifies all the regulatory interactions among genes even with noisy time-series data.

The proposed method was applied in the reconstruction of well-known SOS DNA repair system in *Escherichia coli*. Among 40 genes of SOS network, 6 genes have been considered in this work which controls the core repair system (Little et al., 2013). The expression values of this gene network are measured in a 50-step time-series, and documented in Uri Alon Lab.

(<http://www.weizmann.ac.il/mcb/UriAlon/>)

The experimental result represents biological plausibility of the estimated GRN, which has been validated from various aspects, ranging from the activity of functionally coherent gene sets, to previous experimentally verified interactions among genes.

The rest of the paper is organized as follows. The next section explains the RNN model for reconstructing gene regulatory network followed by the description of the fitness functions used in the proposed methodology. Then, elitist DEMO algorithm for inferring RNN model based GRN has been described which is followed by the section presents the experimental results to highlight the effectiveness of the proposed method. The final section concludes the paper with some general discussions.

## RECURRENT NEURAL NETWORK (RNN) MODEL FOR GENE REGULATORY NETWORK

The Recurrent Neural Network (RNN) model offers a good compromise between the biological proximity and mathematical flexibility while reconstructing gene regulatory network. The model formulates the interactions among the genes in terms of a tightly coupled system (Noman et al., 2013, Vohradsky J, 2013, Wahde and Hertz, 2013) expressed as,

$$z_i = \frac{1}{\tau_i} \left( g \left( \sum_{j=1}^N w_{ij} e_j + \beta_i \right) - \lambda_i e_i \right) \quad (1)$$

where  $N$  ( $i, j \leq N$ ) is the total number of genes in the network,  $e_i$  represents the total regulatory input for  $i$ -th gene,  $w_{ij}$  represents the type and strength of the regulatory interaction of gene- $j$  on gene- $i$  which is either positive (activation), negative (repression) or zero (no regulation).  $\beta_i$  denotes the basal expression level and  $\lambda_i$  represents the decay rate parameter of gene- $i$ . The non-linearity of GRN is introduced by the function  $g()$  which is often given by the sigmoid function. The reconstruction of a gene network using RNN model can be described by the set of parameters  $\{w_{ij}, \beta_i, \lambda_i, \tau_i\}$  which can mimic the experimentally observed gene expression data.

For biological realism, the expression level of gene- $i$  at time  $t + 1$ , i.e.  $e_i(t+1)$  is obtained by normalizing  $z_i$  using a sigmoid squashing function:

$$e_i(t+1) = \frac{1}{1 + e^{-z_i(t)}} \quad (2)$$

The dynamic interactions among genes of a network are reflected in the change of the magnitude of parameter  $w_{ij}$ . In particular, increased values of  $w_{ij}$  indicate strengthened interaction between gene  $i$  and  $j$ , and decreased values indicate weakened interaction. Another important point is that the interactions between the genes have been modeled in this paper based on their expression levels which is a common choice for many existing methods (Noman et al., 2013, Noman and Iba, 2007, Kabir et al., 2013, Ahmed et al., 2013).

## MODEL EVALUATION CRITERIA

The large parameter set of recurrent neural network model emerges the needs of some assessment mechanisms for evaluating the alternate gene networks that come across in the course of evolutionary process. The most commonly used model evaluation process known as *Mean Squared Error (MSE)* is the quantitative difference between the response generated by the candidate model and the experimentally collected response. The smaller the value of MSE, the better the match between observed and calculated expression dynamics, the better the approximation. Like other dynamic systems, the reverse engineering of GRN achieves higher accuracy if multiple time series for the same gene is used. Using  $M$  sets of time dynamics, the MSE based fitness function can be given by

$$f_1 = \sum_{k=1}^M \sum_{t=1}^T \sum_{i=1}^N \left( \frac{e_{k,i}^{cal}(t) - e_{k,i}^{exp}(t)}{e_{k,i}^{exp}(t)} \right)^2 \quad (3)$$

Here,  $e_{k,i}^{exp}(t)$  represents the experimentally observed expression level of gene- $i$  at time  $t$  in the  $k^{th}$  data set. Whereas,  $e_{k,i}^{cal}(t)$  is the numerically calculated expression level of gene- $i$ , at sampling time  $t$  in the same data set which is acquired by solving Equations (1) and (2). Here  $M$  is the number of experimental data sets used,  $T$  is the number of sampling time points and  $N$  represents the number of genes in the regulatory system.

In a biological system very few genes or proteins interact with a particular gene. Because of the high degree of freedom of the model, there exist many local minima in the search space that can also mimic the time courses very closely. Therefore, if all possible regulations are allowed, the search algorithm may get stuck on some locally optimum solution and fail to obtain the true skeletal network structure. To overcome this problem another fitness function is used similar to that used in (Ahmed et al., 2013) as the second objective in the proposed multi-objective inference algorithm. The value of this fitness function is calculated by summing up the number of regulatory inputs of all

the genes in the system. The smaller the value of this fitness function, the sparser the underlying skeletal network structure, closer approximation of the biological reality. Thus for each set of parameters representing regulation networks in recurrent neural network system, the fitness function for obtaining globally optimal gene network structure is given by

$$f_2 = \sum_{i=1}^N I_i \quad (4)$$

Here,  $I_i$  is the number of regulatory inputs to gene- $i$  and  $N$  is the number of genes in the regulatory system.

### PROPOSED INFERENCE METHOD

In this work, an enhanced *Multi-Objective Evolutionary Algorithm (MOEA)* has been used to estimate the model parameters for the target gene regulatory network. Elitist version of DEMO is used in the core of our algorithm as the optimizer that minimizes both  $f_1, f_2$  given in equations (3) and (4) respectively. Like most of the MOEAs, DEMO is a population-based search heuristic, where each individual of the population represents a candidate solution of the problem under consideration. A feasible solution,  $x$  dominates another feasible solution  $y$ , if and only if  $x$  is better than  $y$  for at least one objective function value. An optimum solution called *Pareto optimum* is the one which is not dominated by any other solution in the search spaces. In MOEA, it is impossible to improve a *Pareto optimum* solution with respect to any objective without worsening at least another objective (Storn and Price, 2013, Robic and Filipic, 2013). After random initialization of first generation, each successive new generation is created as follows:

Let,  $P_t$  is the current generation and  $NP$  represents the population size. A new offspring population  $Q_t$  of size  $NP$  is generated by using crossover and mutation operator of DE (Storn and Price, 2013). Each individual of  $Q_t$  can be a newly generated offspring or it can come from  $P_t$ , based on the following principles (Robic and Filipic, 2013).

- 1) The candidate replaces the parent if it dominates it.
- 2) If the parent dominates the candidate, the candidate is discarded.
- 3) Otherwise (when the candidate and parent are non-dominated with regard to each other), the candidate is added to the population.

A combined population,  $R_t = P_t \cup Q_t$  of size between  $NP$  and  $2 \times NP$ , is generated and each individual of  $R_t$  is evaluated using equations (3) and (4). If the population has been enlarged, it is truncated to prepare for the next step of the algorithm.

The truncation process used in this paper is based on NSGA-II (Deb et al., 2013) and comprised of two steps. First, the fast non-dominated sorting (Deb et al., 2013) is applied and individuals of  $R_t$  are sorted into non-dominated fronts  $F_0 \dots F_b$ , where the best non-dominated solutions are stored in  $F_0$ . The members of one front are non-dominated by each other. The next generation,  $P_{t+1}$  is filled, firstly with the members of  $F_0$  and subsequently adding the members of following fronts. However, all members of a front may not be added, because otherwise  $NP$  (number of population) would be exceeded. In this case, crowding distance (Deb et al., 2013) is used to identify diverge non-dominated solutions which will be forwarded to next generation.

Because of the high-degree of freedom of the RNN model, the search space contains many local optimums which may trap the search algorithm and global optimum may remain undiscovered (Noman et al., 2013). Thus, if the fitness values (i.e.,  $f_1$  and  $f_2$ ) of the best compromised individual does not improve for  $G_m$  consecutive generations, the mutation operation of DEMO is evoked which mutates all the other individuals in the current generation. The  $w, \beta$  and  $\tau$  parameters of an individual are mutated by adding random numbers drawn from Gaussian distribution with mean  $\mu_r = 0$  and standard deviation  $\sigma_w, \sigma_\beta$  and  $\sigma_\tau$ , respectively. The  $\lambda$  parameter is mutated using

random numbers drawn from a distribution with mean  $\mu_a = 0$  and standard deviation  $\sigma_a$ .

After the random start, the algorithm proceeds in its regular mode- repeating the above process for all genes until the termination criterion is not met. The output generated by any MOEA is the non-dominated set of solutions known as the *Pareto-optimal* solutions (Robic and Filipic, 2013, Deb et al., 2013). However the decision maker may have imprecise or fuzzy goals for each objective function. Thus, upon having the Pareto-optimal set, a fuzzy based mechanism described in (Abido M, 2013), has been incorporated in the proposed methodology to extract a Pareto-optimal solution as the best compromise solution.

## EXPERIMENTAL RESULTS

The suitability of the proposed GRN reconstruction methodology using RNN model has been primarily validated using a synthetic network as the actual structure and parameter values are unknown for real networks. The experiments were carried out under the ideal noise-free condition and with simulated noise corrupted gene expression data. Finally, the proposed methodology was applied in the reconstruction of SOS DNA repair system of *Echericha coli* using real micro array data.

### Artificial Network Inference

At first, this paper investigated whether it is possible to infer the regulatory interactions and correct parameter values for a small scale 5 gene synthetic network that is also studied by (Ahmed et al., 2013). The regulatory weight matrix of this five genes network is shown in Table 1. The network contains both positive and negative regulations along with feedback loop. The initial gene expression level was selected randomly. In order to simulate the noise experienced in real gene expression data, expression profiles have been generated by adding 5% and 10% Gaussian noise. The experiments were conducted for each condition using 10 sets of data where search ranges for RNN parameters were set as follows

:  $w_{ij} \in [-10.0, 10.8]$ ,  $\beta_i \in [-10.0, 10]$ ,  $\tau_i \in [0.0, 20.0]$ . In the inference of this small scale synthetic network,  $\lambda_i = 1$  is used for all genes. Thus  $\lambda_i$  was not been included in the search as it is fixed for the target. The algorithm was implemented in Java and experiments were run in a Intel(R) Core(TM)2 Duo 2.80 GHz, 2GB RAM - personal computer. Each experiment has been repeated 10 times to confirm the reliability of the proposed GRN reconstruction methodology. This approach ensures that even if the significant solutions of one run miss a true regulation, the subsequent runs may find that. That is, the outputs from all of these run are taken into consideration for ensuring the validity of the algorithm.

**Table 1.** Weight Matrix for target synthetic network

Gene	1	2	3	4	5
1	-1.30	0.0	2.86	0.0	-0.70
2	0.80	-1.27	0.0	0.0	0.0
3	0.0	-0.86	-1.70	0.0	0.0
4	0.0	0.0	1.66	-1.37	-0.70
5	0.0	0.0	0.0	1.70	-1.70

**Table 2.** Inferred Weight Matrix for target synthetic network using 5% noisy time-series data

Gene	1	2	3	4	5
1	-1.29	0.0	3.00	0.20	-0.77
2	0.85	-1.40	0.0	0.0	0.0
3	0.0	-0.78	-1.71	0.0	-0.01
4	0.0	0.0	0.98	-1.65	-0.55
5	0.0	0.0	0.0	1.57	-1.57

**Table 3.** Inferred Weight Matrix for target synthetic network using 10% noisy time-series data

Gene	1	2	3	4	5
1	-1.29	-0.11	2.30	-0.40	-0.47
2	0.72	-1.30	0.18	-0.36	0.0
3	-0.08	-0.81	-1.72	-0.33	0.0
4	0.0	0.28	1.34	-1.14	-0.77
5	-0.13	0.0	0.0	1.22	-1.31

**Table 4.** Average SN, SP of the target network for noise-free, 5% and 10% noisy time-series data

	SN	SP
Noise-free	1.00	1.00
5% Noisy	1.00	0.60
10% Noisy	1.00	0.55

In almost every optimization run with noise-free expression data, fitness score for models reach to zero or very close to zero ( $< 10^{-6}$ ) and the estimated parameters are exactly the same as the target. The performance of the reconstruction algorithm is also analyzed using noisy time-series data with the same experimental conditions. Table 2 and 3 shows the estimated network structure and parameter values achieved in a sample run for 5% and 10% noisy data respectively. From Table 2 and 3, it is evident that even in the presence of high level of noise the proposed method has successfully predicted all the regulatory interactions among the genes. Some false positive regulations are also predicted by the search algorithm while working with noisy data. However, the magnitudes of these false positives were pretty small compared to the real regulations. The summary of prediction in terms of *sensitivity* (SN) and *specificity* (SP) has been presented in Table 4 using their standard definition based on positive/negative value of  $w_{ij}$ . This result shows that the prediction contains a full 1.00 sensitivity and the specificity greater than 0.50 even for corrupted GRN data. In the case of 10% noisy data the specificity value 0.55 means prediction of 45% false positive regulations. In an overall, the proposed approach performs a correct prediction of the network structure and a good approximation of the model parameters.

### Analysis of Real Microarray Data

The proposed methodology has been analyzed in the reconstruction of well-known SOS DNA repair network in *Escherichia coli*. It is the longest, most complex and best understood DNA damage-inducible network to be characterized to date. In this work, the experiment was carried out by the gene expression data set collected in Uri Alon Lab.

The data set contains expression levels of 8 genes namely *uvrD*, *lexA*, *umuD*, *recA*, *uvrA*, *uvrY*, *ruvA*, *polB*. Four experiments were done using various light intensities, in each of which 50 samples were collected at 6 minutes interval for the above 8 genes (Perrin et al., 2013). For reconstructing GRN, this paper used the data sets from experiment 3 and 4. To meet biological reality, data corresponding to each gene was normalized within the range (0, 1] against their maximum value and very small value ( $\sim 10^{-4}$ ) was used to replace all the zero expression levels in these two data sets.

In this work, 6 key regulators namely *uvrD*, *lexA*, *umuD*, *recA*, *uvrA* and *polB* have been considered in the reconstruction process. This sub network is well studied one and the interactions among different genes are known. Being actual microarray data, there is unknown amount of noise inherently present in these data. These noises in the data may have had an influence on the inference method. So, the generated results have been much dispersed. The results have been generated based on the different runs of the algorithm.

The regulations of each gene have been identified using the following search ranges of RNN parameter:

$w_{ij} \in [-10.0, 10.0]$ ,  $\beta_i \in [-10.0, 10.0]$ ,  $\tau_i \in [0.0, 15.0]$  and  $\lambda_i \in [0.0, 1.0]$ . The known regulations and the predicted regulations for all the 6 genes in the SOS repair network identified by the proposed algorithm have been summarized in Table 5.

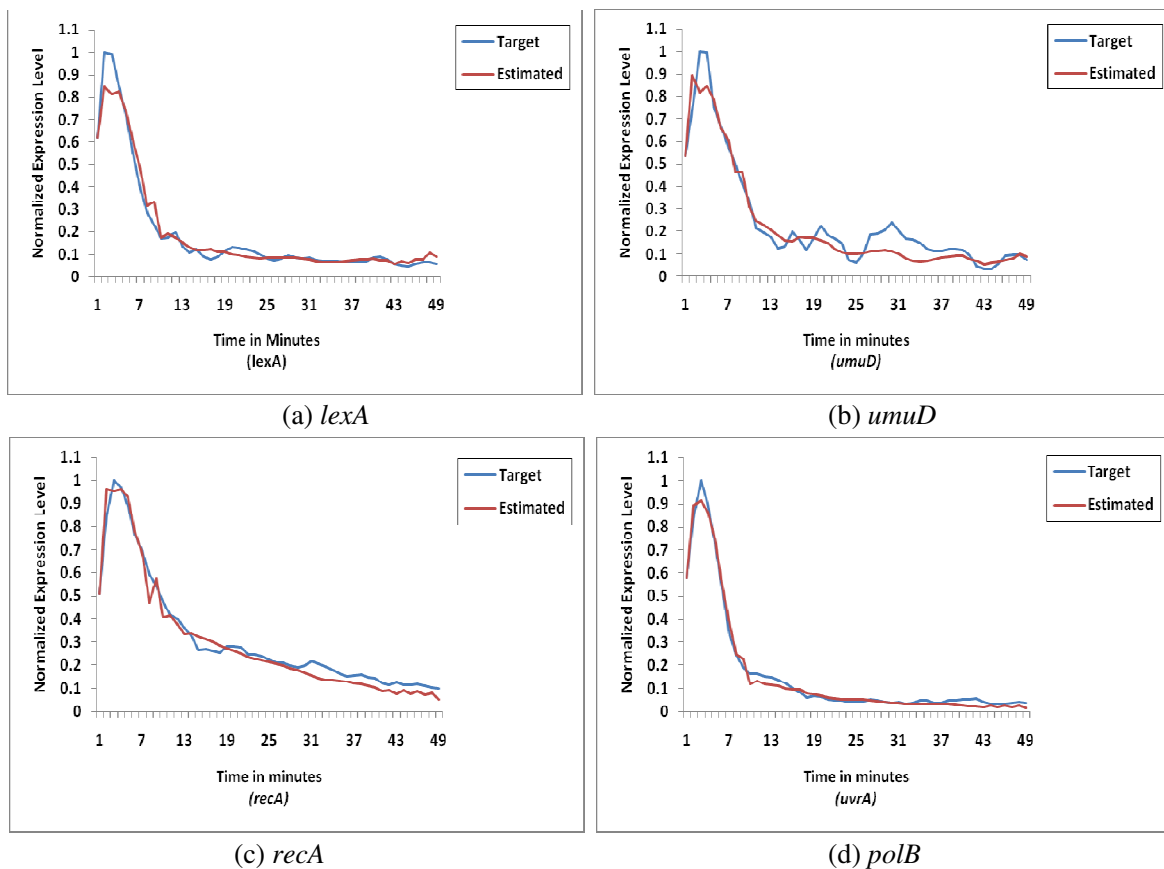
In each run, the reconstruction process achieves a very small fitness function value which indicates that the inferred network model could match the target time course data pretty well. The comparison between the target dynamics and the estimated model generated dynamics for some selected genes has been shown in Figure 1. From Figure 1, it is evident that the proposed method has the ability to mimic the system dynamics adequately.

**Table 5.** Estimated regulations for SOS DNA repair system

	<i>uvrD</i>	<i>lexA</i>	<i>umuD</i>	<i>recA</i>	<i>uvrA</i>	<i>polB</i>	References
<i>uvrD</i>	-	-					(Shuheï et al., 2013, Cho et al., 2013, Shuheï et al., 2013)
<i>lexA</i>	+	+	+	-			(Shuheï et al., 2013, Cho et al., 2013, Shuheï et al., 2013)
<i>umuD</i>		-	+	-			(Noman et al., 2013, Shuheï et al., 2013, Bansal et al., 2013, Gardner et al., 2013)
<i>recA</i>		-	+	+		+	(Shuheï et al., 2013, Bansal et al., 2013)
<i>uvrA</i>		+		-	-		(Kabir et al., 2013, Shuheï et al., 2013, Perrin et al., 2013)
<i>polB</i>		-			+		(Noman et al., 2013, Kabir et al., 2013, Shuheï et al., 2013)

The estimated regulations and parameter values were different from run to run in the conducted experiments. However, examining all of the extracted interactions with regard to known roles of selected genes; it is evident that, in most cases, the predictions confirm the prior knowledge, which

indicates the suitability of the proposed method. The algorithm also predicts a number of false positives which are either unknown regulations or the side effect of noise presented in micro array data.



**Figure 1.** Target and estimated dynamics for the SOS DNA repair system



## DISCUSSION AND CONCLUSION

Gene regulatory networks are abstract mapping of the more complicated biochemical systems and inherently nonlinear in nature. The inference of the large scale GRN is always impeded by the computational requirements imposed by the underlying model. In this work, recurrent neural network model is used to infer the target gene expression profiles and found very effective in terms of biological actuality and computational feasibility. However, the RNN model contains a large number of parameters and because of the high-degree of freedom of the model; the search becomes very complicated and the global optimum solution may remain undiscovered. To overcome this problem, a second objective function based on skeletal network architecture, has been incorporated in the proposed method which ensures the inference of sparser biological networks. A natural multi-objective computational approach, known as DEMO is used to infer the true structure of underlying biological system. Among the EA based multi-objective search heuristics, elitist version of DEMO is used in the proposed methodology because of its reputation of fast convergence in complex and conflicting search spaces.

Some experimental analysis of the proposed method has been performed to investigate the different components of the algorithm which are necessary for accurate estimation of the regulatory parameters. All of the results are based on experimenting with an artificial gene network and analyzing a real micro array gene expression profile. The performance of a reverse-engineering algorithm always affected by the noise levels presented in the experimental data and the proposed methodology is no exception. Thus, the synthetic gene expression data corrupted with varying noise levels have been used to highlight practicability of the proposed optimization algorithm in estimating robust parameter values. From the experimental results, it is very evident that the proposed method is very efficient in the estimation of true network structure even in the

presence of high levels of noise. Moreover, the two performance measures, i.e. SN and SP, showed the resistance of the proposed approach in the case of identifying the false regulations among genes. In the analysis of SOS DNA repair network of *E. coli*, because of the insufficient amount of gene expression data with high noise, it was very difficult for the proposed method to get any consistent result for the target network in the different experimental runs. Nonetheless, most of the pathways in the reconstructed network were consistent with the results reported in the literature. Although, the proposed reverse-engineering algorithm may not be able to capture the complete network architecture in a single run, because of insufficient data availability corrupted with excessive noise; still, this type of indication can be very useful for the biologists to design additional experiments that may in turn help to identify new interactions among the genes.

With the speedy growth of biological samples categorization and characterization, and enhanced data collection techniques, it is expected that high-dimensional and feature-rich data will be collected which will represent complex dynamics of biological systems. Thus, the development of decoupled version of the proposed method will be a timely contribution to narrow the gap between the imminent methodological needs and the available biological data. Moreover, such decoupling of the original method not only offers a deeper understanding of the mechanisms and processes underlying biological networks, but also eases the immediate parallelization or distributed implementation of the proposed GRN reconstruction algorithm.

## REFERENCES

1. Noman N, Palafox L, Iba H. Reconstruction of gene regulatory networks from gene expression data using decoupled recurrent neural network model. In *Natural Computing and Beyond*, Springer Japan. 2013, 93-103.
2. Noman N, Iba H. Inferring gene regulatory networks using differential evolution with

- local search heuristics. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*. 2007, 4(4): 634-647.
3. Schena M. DNA microarrays: a practical approach. 1999.
  4. Sahoo, Debashis, David L. Dill, Andrew J. Gentles, Robert Tibshirani, Sylvia K. Plevritis. Boolean implication networks derived from large scale, whole genome microarray datasets. *Genome Biology*. 2008, 9(10): R157.
  5. Dhaeseleer P, Wen X, Fuhrman S, Somogyi R, Linear modeling of mRNA expression levels during CNS development and injury. *Pacific Symposium on Biocomputing*, 1999, 4:41-52.
  6. Mazur J, Kaderali L. The importance and challenges of Bayesian parameter learning in systems biology. *Model Based Parameter Estimation*, Springer Berlin Heidelberg. 2013, 145-156.
  7. Vohradsky J. Neural model of the genetic network. *Journal of Biological Chemistry*. 2001, 276(39): 36168-36173.
  8. Chen T, He H, Church G. Modeling gene expression with differential equations. *Pacific Symposium on Biocomputing*. 1999, 4: 29 - 40.
  9. Kabir M, Noman N, Iba H. Reverse engineering gene regulatory network from microarray data using linear time-variant model. *BMC Bioinformatics*. 2010, 11(1): S56.
  10. Savageau M A. *Biochemical Systems Analysis. A Study of Function and Design in Molecular Biology*. Addison-Wesley. 1976.
  11. McAdams H, Arkin A. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences*. 1997, 94(3): 814.
  12. Chowdhury A R, Chetty M, Vinh N X. Incorporating time-delays in S-System model for reverse engineering genetic networks. *BMC bioinformatics*. 2013, 14(1): 196.
  13. Basso K, Margolin A A, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. *Nature genetics*. 2005, 37(4): 382-390.
  14. Wahde M, Hertz J. Coarse-grained reverse engineering of genetic regulatory networks. *Biosystems*. 2000, 55(1): 129-136.
  15. Chiang JH, Chao SY. Modeling human cancer-related regulatory modules by GA-RNN hybrid algorithms. *BMC Bioinformatics*. 2007, 8(1): 91.
  16. Ahmed S, Hasan M M, Noman N. Reconstructing Gene Regulatory Network using Linear Time-Variant Model. *Dhaka University Journal of Applied Science and Engineering*. 2011, 1(2): 125-129.
  17. Sultana R, Showkat D, Samiullah M, Chowdhury A R. Reconstructing Gene Regulatory Network with Enhanced Particle Swarm Optimization. *Neural Information Processing*, Springer International Publishing. 2014, 229-236.
  18. Little J W, Edmiston S H, Pacelli L Z, Mount D W. Cleavage of the Escherichia coli lexA protein by the recA protease. *Proceedings of National Academy of Science (PNAS)*. 1980, 77(6): 3225-3229.
  19. Storn R, Price K. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*. 1997, 11(4): 341 - 359.
  20. Robic T, Filipic B. DEMO: Differential evolution for multiobjective optimization. In *Evolutionary Multi-Criterion Optimization, Third International Conference, EMO*, Springer Berlin Heidelberg. 2005, 520-533.
  21. Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE TRANS EVOL COMPUT*. 2002, 6(2): 182 - 197.
  22. Abido M. Multiobjective evolutionary algorithms for electric power dispatch problem. *IEEE transactions on evolutionary computation*. 2006, 10(3): 315 - 329.
  23. Perrin B E, Ralaivola L, Mazurie A, Bottani S, Mallet J, d'Alch'e-Buc F. Gene networks inference using dynamic bayesian networks. *Bioinformatics*. 2003, 19: 138-148.

24. Shuhei K, Katsuki S, Soichiro Y, Hideki M, Koki M, Mariko H. Function approximation approach to the inference of reduced NGnet models of genetic networks. *BMC Bioinformatics*. 2008, 9(1): 23.
25. Cho D, Cho K, Zhang B. Identification of biochemical networks by S-tree based genetic programming. *Bioinformatics*. 2006, 22(13): 1631.
26. Shuhei K, Satoshi N, Mariko H. Genetic network inference as a series of discrimination tasks. *Bioinformatics*. 2009, 25(7): 918-925.
27. Bansal M, Gatta G, Bernardo D D. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics*. 2006, 22(7): 815.
28. Gardner T, Bernardo D D, Lorenz D, Collins J. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, 2003, 301(5629): 102.
29. Perrin B, Ralaivola L, Mazurie A, Bottani S, Mallet J, d'Alche Buc F. Gene networks inference using dynamic Bayesian networks. *Bioinformatics*. 2003, 19(2): 138 – 148.