Journal of Advanced Bioinformatics Applications and Research
ISSN 0976-2604.Online ISSN 2278–6007
Vol 5, Issue3, 2014, pp121-133
http://www.bipublication.com

Two Weight Matrix Model Based GNET Reconstruction Using Clonal Algorithm

Jereesh A S[*] and Govindan V K

Department of Computer Science & Engineering,
National Institute of Technology, Calicut, Kerala, India
[*]Corresponding author: Email:jereesh.a.s@gmail.com

ABSTRACT:

In a particular biological process, gene regulatory network describes how each gene is controlled by a set of genes. Identifying the gene regulatory network through reverse engineering is a widely used approach. In this article, gene regulatory network reconstruction using evolutionary algorithm from microarray gene expression data is attempted. A two-weight matrix model, which is capable of identifying relationships between genes and the mysterious effects for each gene, is proposed. An evolutionary algorithm based on the concepts of artificial immune system called Clonal selection based algorithm is made use of to learn the parameters of the proposed model. A well-accepted error criterion, the mean squared error is used as the fitness function for the method. The proposed approach is tested on an artificial microarray data and a real life data called SOS DNA repair system of E.coli. It is observed that the proposal is capable of finding the major interactions between genes. The convergence of the proposed approach is much faster than many of the existing approaches. The proposed approach also identified some novel regulatory relations that can be useful to the biologists for discovering new facts.

Keywords: Microarray analysis, evolutionary Algorithm, Artificial Immune System, S-system, Gene Regulatory Network, SOS E coli. DNA repairing.

[I] INTRODUCTION

Living world is characterized by the staggering complexity of the systems involved. Intense research is going on around the world to untangle the intricacies underlying these systems. Modeling and simulation of biological entities and their mutual interactions are vital to proper understanding of the systems. Over the years, we have developed many tools and techniques for modeling and understanding of biological systems. Molecular biology is one of the most popular and fundamental approaches in deciphering biological processes. With the help of gene expression data, we can identify a set of genes involving and interacting in a particular biological process. The interactions between different genes for a particular biological process can be modeled as gene regulatory network. In this model, genes are the nodes and the links between them represent the relationship between genes for a particular condition of a biological process. With the help of proper gene regulatory network modeling it is possible to identify how each of the genes is participating in a particular biological process. Drug design is one of the important applications, which makes use of gene regulatory network. Identification of the set of

genes that participate in a biological process allows us to develop drugs that can directly affect those genes for the improvement of biological conditions.

An evolutionary algorithm is a computational paradigm, which derives its cues from the quintessence of biological evolution. Biological evolution is marked by a progressive improvement in the overall quality of the population undergoing the process of evolution. In the same way, evolutionary algorithms attempt to solve a problem starting with a random population of candidate solutions, which undergo a process similar to organic evolution wherein each generation of candidates is qualitatively better than the previous generations. The process is driven by an equivalent of random mutation and natural selection where the less desirable candidates are annihilated from the population. Evolutionary algorithms are powerful optimization tools and are very popular in the solution of gene-regulatory networks. There are many evolutionary algorithms used for the modeling of gene regulatory network. One of the major goals of reconstruction of gene regulatory network, apart from accuracy, is the reduction of number of fitness value calculations used to arrive at the convergence of the error in the optimization process. The proposal in this paper attempts faster convergence of the optimization process.

The problem of gene regulatory network reconstruction mainly contains two major aspects: identifying a mathematical model for the gene regulatory network and learning the parameters for the model. Most of the research reported concentrate mainly on one of the above aspects. Objective of this paper is to propose a new approach, which leads to improvements in processing time and accuracy.

The rest of the paper is organized as follows: Section 2 presents the relevant reviews on various techniques employed for the gene regulatory network reconstruction. The proposed mathematical model and algorithm for the optimization process is illustrated in Section 3. Section 4 describes the experimental setup and comparison of the results of the new proposal with other methods. Section 5 discusses the performance of the proposal on real life data, and finally the paper is concluded in Section 6.

[II] LITERATURE SURVEY

Researchers have proposed a number of evolutionary algorithms for the construction of gene regulatory networks. Ando et al. [1] has proposed a LMS-GP algorithm that uses Genetic programming (GP) to reduce mean-square-error between observed and experimentally identified arrays. In this algorithm, a general form of differential equation is used to model the system. According to this algorithm, for each generation, the individuals are generated as tree structures for recombination, which is a complex process. Another method is the Genetic Program proposed by Wang et al. [2] which makes use of Kalman filter for estimating the parameters of the model. In this, GP predicts the structure of the model. The algorithm requires the noise statistics for the successful optimization of the parameters of the regulatory network using Kalman filtering. Hence, it is not easy to apply for the reconstruction of GRN.

Noman et al. [3] proposed a method that consists of a decomposed S-system model and an extended version of the fitness function. S-system is a power full nonlinear model proposed by Savageau [4] based on the mathematical modeling of chemical processes. An evolutionary algorithm called trigonometric differential evolution along with a greedy search is employed to optimize the parameters. The effect of decoupled S-system reduces the dimensionality in computation. Chowdhury and Chetty [5] extended the work of Noman et al. [3] by introducing the following concepts: prediction initialization (PI), Flip operation (FO) and a restricted Hill climbing search for selected

individuals. A major disadvantage of the approach is the reduction of accuracy.

A software, named Gene Network Explore (GNetXP) developed by Chan et al. [6] is employed for the reconstruction of Gene regulatory network. In this, the gene profile is clustered and the ordinary differential equation parameters are optimized using Kalman filter. For clustering purpose, a hybrid algorithm that consists of Genetic algorithm and expectation maximization algorithm is employed. Another work is the adaptive fuzzy evolutionary gene regulatory network reconstruction framework (AFEGRN) proposed by the Shoaib et al. [7]. This approach is based on the fuzzy clustering using EA and Spearman correlation.

In [8], Ram and Chetty made use of guiding operators for a speedy result. Guided GA also concentrates on diversity of the individuals in each generation in order to avoid the local optimum convergence. This algorithm uses two levels of knowledge discovery, leading to increased complexity.

Tominaga and Horton [9] considered biological network as a scale free network and used advancement of GA called Distributed genetic algorithm to optimize S-system parameters. The properties of biological systems as a prerequisite are necessary for the domain knowledge of network. The issue here is that the knowledge about such properties are often not available.

An intelligent two stage evolutionary algorithm (iTEA) proposed by Ho et al. [10] is a divide-and-conquer method of identifying gene network. In this the authors introduced an intelligent Genetic Algorithm (IGA) for solving the sub problems which was decomposed initially and then orthogonal experimental design based simulated annealing (OSA) was used to combine and extract the overall result. An improvement for iTEA, called iTEA2 proposed by Shu et al. [11], employed a novel chromosome encoding and crossover method. The approach makes use of domain knowledge for the processing of GRN.

Maraziotis et al. [12] developed a hybrid method which combines the dynamic fuzzy rules and neural network called multilayer evolutionary trained neuro-fuzzy recurrent network (ENFRN). In ENFRN, fuzzy rules are time dependent and are updated on each time periods. After the first level of the algorithm, ENFRN makes use of the Particle Swarm Optimization algorithm for the optimization of the structure.

An evolutionary bi-clustering method proposed by Mitra et al. [13] tried to extract the bi-clusters using multi objective evolutionary algorithm (MOEA) in order to reduce the dimensional complexity. Correlation matrices are used for the identification of interactions. The disadvantage is that there can be delay for effecting the changed TF to target, leading to low correlation metric.

A multi-objective evolutionary optimization algorithm proposed by Schroder et al. [14] used GA for optimizing three objective functions. First function is to cluster genes in such a way that maximizes the common patterns in the transcription factor. Second function concentrates on the regulatory relationships between genes and the third objective function concentrates on maximizing the pathway score. As there are three objective functions involved, the system will be complex.

Gene Regulatory Network Modeling using Cuckoo Search and S-system [15] used a cuckoo search method for the optimization of the S-system. This approach converges at a faster rate when compared to existing Clonal selection based algorithm using S-system [16].

Spieth et al. [17] proposed a memetic inference method for gene regulatory network based on S-system. The memetic algorithm is a hybrid algorithm, which employs a combination of genetic algorithm (GA) and covariance matrix evolution strategy (CMES) [17]. GA is used to optimize the structural topology, and the evolutionary strategy is a local search algorithm for optimizing the S-system parameters. This is considered as a standard algorithm, which is used

for comparative studies of new proposal of this paper.

S-system is the most well accepted and standard differential equation model introduced by Savageau [4]. Even though S-system is the best model as per the current state of art, this model has disadvantages. Number of parameters in the model is large and it will reduce the convergence speed for the problem. The exponential terms will again affect the computational speed. In order to avoid such disadvantages in this paper introduced a new model called Two Weight Matrix model (TWM).

[III] PROPOSED METHOD

The proposed approach combines a new model called Two Weight Matrix model with the Clonal selection based algorithm. Clonal selection is an algorithm emulating the clonal selection process in vertebrate immune system to create an evolving population of candidate solutions, which culminate in an optimal solution to the problem. These are briefly introduced in the following subsections.

3.1. TWM Model

Gene Regulatory Network (GRN) is an inter-connected network of N number of genes. Each gene g_i is expressed as x_i with respect to other genes, under a particular circumstance. Therefore, we can represent the change of expression concentration over time as, x (t+1) = h(X(t)) where X(t)=(x_1,….,x_N), t is the current time and h() is the operation corresponding to the gene regulatory network behavior.

To reconstruct the gene regulatory network a new approach called Two-Weight-Matrix (TWM) model, given below, inspired from S-system model is introduced.

$$\frac{dx_i}{dt} = \sum_{j=1}^{N} w_{ex_{i,j}} * x_j(t) - \sum_{j=1}^{N} w_{in_{i,j}} * x_j(t) + \mathfrak{E}_i \qquad (1)$$

The Two-Weight-matrix model consists of two weight matrices and an epigenetic factor. Excitatory and inhibitory weight matrices are used for the representation of the relationships/ interactions between genes. The epigenetic factor ($\mathfrak{E}$) represents all other external effects that leads to the regulation of genes. Excitatory weight matrix (Wex) values up-regulate the gene expression rate, and the inhibitory weight matrix (Win) values down-regulate the gene expression rate. Wex and Win determine the structure of the regulatory network. Compared to S-system model this model has less number of variables, that is, $2*N^2+N$, and has higher speed as the model avoided the exponential terms of S-system.

3.2. Artificial Immune System (AIS)

Artificial immune system is derived from the developments in immunology. Theoretical immunology explains the structuring and functioning of the immune system. Natural Immune Mechanism has the typical characteristics of a robust pattern recognition system. It is adaptive, robust to slight variations in the nature of antigens recognized and evolves progressively in terms of its capability to identify pathogens. The process of antigen recognition can be emulated to create a pattern recognition system that also evolves and improves over generations of candidate prototypes. Each time, through learning, the immune system increases its durability. The learning strength and memorizing property of immune system processes can be made use of in solving the optimization problems. Castro and Zuben [18] identified a set of important characteristics, namely, pathogen recognition, distributed way of detection, and noise tolerance. In this paper, the model of the Gene regulatory network is optimized using the artificial immune system technique. The algorithm makes use of the adaptive immunity mechanism [19] called clonal based selection [20].

3.3. CLONAL SELECTION ALGORITHM

Burnet [21] illustrated the fundamental theory of immune system with the help of a mechanism

called clonal selection. The introduction of Burnet's theory led to a new era in biological immune system. Based on these concepts a widely used clonal selection based algorithm called CLONALG [22] was developed by Castro and Zuben. In the clonal selection algorithm, an optimal antibody is identified through a set of iterative operations such as selection, cloning, affinity maturation and reselection. In order to improve the results, the property of randomness is merged in the CLONALG by introducing random antibodies in each generation. Affinity is a measure of closeness to optimal antibody and is calculated using a fitness value. In this algorithm, instead of crossover and mutation faster procedures called cloning and affinity maturation are employed. In the present work, clonal selection algorithm is used to optimize the proposed novel TWM model. The Clonal selection algorithm is given in Algorithm 1. An individual chromosome involved in the Clonal selection algorithm is shown graphically as Fig 1.

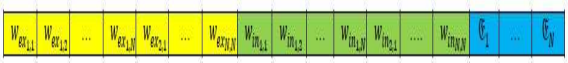

Fig. 1: A chromosome for CLONAL based Algorithm for the optimization

Algorithm 1: CLONAL Selection based Algorithm for the optimization using TWM model

Require: Microarray gene expression data matrix N*T, Maximum Number of Generations G, error tolerance, Number of individuals in a Population P.

Ensure: Optimal antibody

1. Start.

/* Initialize generation number as zero*/

2. Generation := 0

/* Initialize the first generation population*/

3. Pop(Generation) := Init(Clonal pop(P))

/* Fitness evaluation of the current population */

4. Evaluate_Fitness (Pop(Generation))
5. while termination criteria not met do

/*Select individuals from the population pool*/

5.1. Selected_Pop(Generation):=Selection(Pop(Generation))

/*generate Cloned population from the selected individuals*/

5.2. Cloned_Pop(Generation):=Clone(Selected_Pop(Generation)

/*Maturate cloned population and merge with the population pool*/

5.3. Mature(Generation):=Maturation(Cloned_pop(Generation))

/*Randomly generate individuals*/

5.4. Rand(Generation):=Random()

/*Merge and update current population*/

5.5. Pop(Generation)=Merge(Pop(Generation),Mature(Generation), Rand(Generation))

/* Fitness evaluation of the current population*/

5.6. Evaluate_Fitness (Pop(Generation)

/* Select individuals from the population pool for the next generation*/

5.7. Pop(Generation+1):=Re_Selection(Pop(Generation))

5.8. Generation := Generation + 1

6. end while
7. Stop.

The process of optimization is diagrammatically represented as shown in Fig. 2

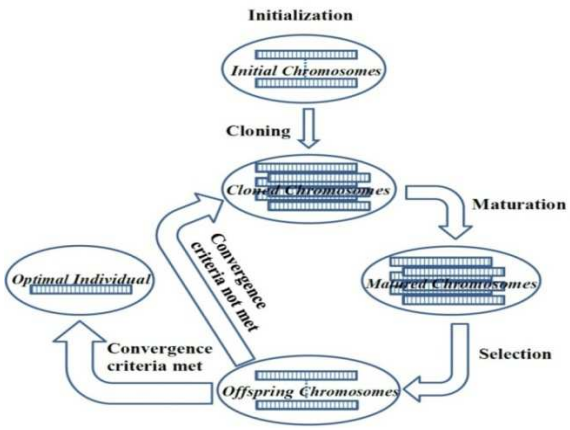

Fig. 2: CLONAL selection based Algorithm for the optimization

Fitness function: The algorithm makes use of following fitness function proposed by Tominaga et al. [23]:

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{X_{i,t}^{cal}-X_{i,t}^{exp}}{X_{i,t}^{exp}}\right)^{2} \quad (2)$$

Where $X^{cal}_{i,t}$, $X^{exp}_{i,t}$ are the expression value of gene i at time t from the estimated (calculated) and experimental data respectively.

Maturate Rate: Maturate rate is the rate at which the selected candidate matured due to weight updating and is defined as follows.

$$Maturate\ rate = 1 * e^{-(Fitness\_best\_pop(generation))} \quad (3)$$

Where *Fitness_best_pop(Generation)* represents fitness value of the best individual of the current generation.

Complexity analysis: An approximate analysis of the complexity of Clonal selection based approach using TWM is given below. The major complexity terms in the algorithm are due to the initialization process, Fitness evaluation, Selection, Cloning, Hyper mutation/maturation and Random individual generation. Based on these, an approximate complexity analysis is carried out in the following:

Initialization:
Initialize Np number of parents of length *L= (2Ng2+Ng)*
Where Np: Number of parent individual,
L: Length of an individual,
Ng: Total number of genes in microarray
Approximate complexity for initialization= *Np*L= (2Ng2+Ng)*Np*

Fitness evaluation:
Step1: Approximate complexity of TWM model proposed (equation 1): *(Ng+Ng+1)*Ng*
Step2: Approximate complexity for differentiating TWM model t times for the calculated microarray= *(2Ng2+Ng)*t.*
Step3: Approximate complexity of mean squared error between calculated microarray and experimental microarray= *Ng*t*
For each *N= (Np+Nc+Nd)*, calculate fitness value in a generation. Hence Approximate fitness evaluation for the fitness evaluation for a generation= *((2Ng^2+Ng)*t+ Ng*t)*N*
Where *Nc*: number of cloned children and
Nd: number of random generated individuals in a generation.

Selection:
Select best *Np*: Number of parents from the pool *(Np+Nc+Nd)*
Sorting using heap sort and select top Np individuals
Approximate complexity for selection= *N*log N* where *N= (Np+Nc+Nd).*

Cloning:
Cloning is an asexual reproduction of parents to form cloned children
Cloning number of each parent = *Nci*=round *((β*Np/i))*
Where *β*: cloned factor; *i*: Rank/position of parent with respect to fitness value
Approximate complexity for cloning=*Nc*

Hyper mutation/maturation:
Mutate each *Abc* (child antibody) with mutation rate of as per equation 3
Approximate complexity for calculating mutation rate =*Nc*.
For mutation, complexity =*Nc*
Approximate complexity for maturation =*Nc+Nc= 2Nc*

Random individual generation:
Approximate complexity of generating *Nd* number of random individuals= *NdL=Nd(2Ng2+Ng)*
Total approximate complexity for the Clonal selection approach is *N*(2Ng2t+logN)Ngeneration*

Since *Ng* and *t* is depends on the input data we can say this algorithm is having *N*(logN)Ngeneration* complexity.

[IV] EXPERIMENTAL SET UP AND RESULTS

For comparing the efficiency of the proposed method, a well-known five gene standard artificial network [3, 5, 9, 10, 24-29] is identified. Table 1 represents the target network model using the S-system parameters. The simulated microarray data is generated using the Runge-kutta algorithm and S-system model [24]. For the experimentation, 10 sets of expression data with initial values in the interval [0,1] are generated artificially. The initial values used for the generation of artificial data are given in Table 2. A sample Time dynamics of the 5-dim regulatory system inferred is shown in Fig.3 where, duration of 0.0 to 0.5 is divided into 11 equi-distance samples, and 10 points are computed between each sampling point.

The performance of the proposed approach is compared with the popular existing memetic method proposed by Spieth et al. [17], which is an efficient method than standard ES and an improved version of GA called skeletalizing [30].

Memetic algorithm: Memetic algorithm proposed by Spieth et al. [17] is a hybrid algorithm, which makes use of Genetic algorithm and CMA-ES. GA is used to optimize the structure of network globally and ES is used to optimize the parameters of the S-system locally. This algorithm uses 3-point cross over with cross over probability and mutation probability as one. In this method, CMA-ES starts with the 10 parents and 50 offspring. Standard S-system is used to model the network using memetic algorithm.

Clonal selection based algorithm using S-system [16]: In order to confirm the efficiency of the Clonal selection based algorithm, the algorithm using S-system model is also implemented. This algorithm has the property of randomness like PSO and evolution as in GA.

Clonal selection based algorithm using TWM model: This method is an improvisation over the clonal selection based algorithm using S-system. This comparison demonstrates the efficiency and speed of TWM model compared to S-system.

All the above three methods are implemented and compared in order to confirm the superior performances of the proposed method using the five gene artificial genetic network [3, 5, 9, 10, 24-29]. As evolutionary algorithms are based on random probability distributions, multiple data sets of similar kinds are used for the analysis of results. Hence, 10 sets of data using the different initial conditions are generated for this purpose. Mean Squared Error (MSE) [30] is used as the error evaluation measurement metric.

The results of the comparison with various methods are given in Figure 4. The memetic method using S-system required 2*106 fitness evaluations for convergence [17] whereas Clonal selection based algorithm using S-system converges at around 3*105 fitness evaluations. This clearly demonstrates the power of Clonal selection based algorithm with compared to memetic algorithm. Clonal selection based algorithm using TWM converges around 1*105 fitness evaluation which is much faster than the first two. Thus, the proposed Clonal selection based TWM algorithm outperforms the memetic and Clonal selection based s-system algorithms.

[V] DISCUSSION

Analysis of real life data using the proposed method

SOS DNA repair system in E.coli.[31] is a famous real life data set which is commonly used to evaluate the efficiency of gene regulatory network reconstruction methods. Figure 5 is a graphical representation of gene interactions after the damage of DNA. Mainly 6 major genes (uvrD, umuD, lexA, recA, uvrA and polB) are involved in the processing of DNA repair [27, 29, 32-35]. LexA is a repressor gene which inhibits the expression of other genes. Whenever a DNA

damage happens in E.coli. RecA identifies the damage and activates the processing of cleavage of LexA. Hence, the concentration of the LexA will be reduced and will lead to the excitation of other genes. After the clearance of damage, cleavage of LexA will be slow downed and stopped, and this leading to increased concentration of LexA. The LexA will repress the other genes and will advance to a balanced state. Construction of gene network allows predicting the roles of each of the genes in the DNA repairing system.

SOS Data is obtained from the experiments done by Uri Alon lab of Weizmann Institute of science from website www.weizmann.ac.il/mcb/UriAlon/Papers/SOSData/. There are 50 time periods for the experiment in which 49 are used for the experimentation where the first time period is at zeroth time and contains zero knowledge. Out of the 8 genes we had selected, 6 important genes are specified. All the values in the expression are normalized in the range of [0, 1].

i	α_i	$G_{i,1}$	$G_{i,2}$	$G_{i,3}$	$G_{i,4}$	$G_{i,5}$	β_i	$H_{i,1}$	$H_{i,2}$	$H_{i,3}$	$H_{i,4}$	$H_{i,5}$
1	5.0	0.0	0.0	1.0	0.0	-1.0	10.0	2.0	0.0	0.0	0.0	0.0
2	10.0	2.0	0.0	0.0	0.0	0.0	10.0	0.0	2.0	0.0	0.0	0.0
3	10.0	0.0	-1.0	0.0	0.0	0.0	10.0	0.0	-1.0	2.0	0.0	0.0
4	8.0	0.0	0.0	2.0	0.0	-1.0	10.0	0.0	0.0	0.0	2.0	0.0
5	10.0	0.0	0.0	0.0	2.0	0.0	10.0	0.0	0.0	0.0	0.0	2.0

Table 1: S-system model parameters for the target network model

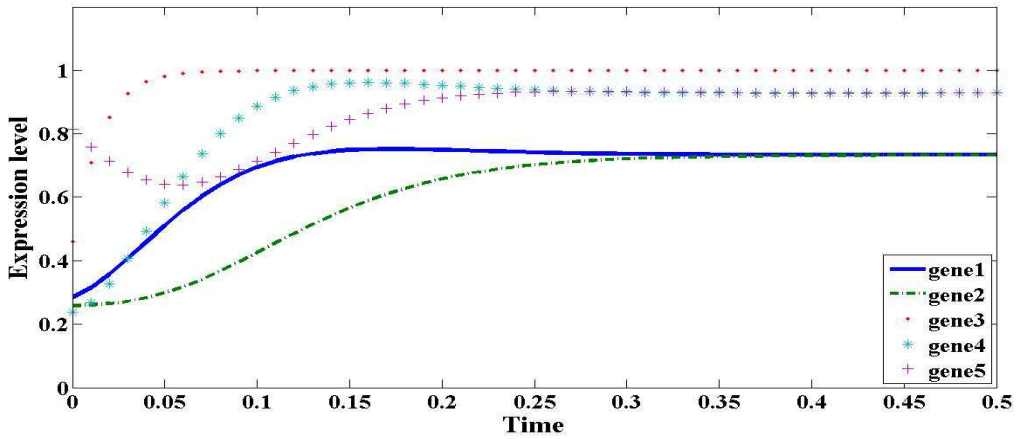

Fig. 3: A sample Time dynamics of the 5 dimensional regulatory system using parameters in Table 1.

Data sets / Genes	Set 1	Set2	Set3	Set4	Set5	Set6	Set7	Set8	Set9	Set10
G1	0.8231	0.2851	0.9961	0.9991	0.7937	0.1479	0.6264	0.9556	0.6724	0.4216
G2	0.3933	0.2586	0.0400	0.0770	0.5441	0.2278	0.9497	0.6866	0.2542	0.6126
G3	0.6273	0.4616	0.5457	0.9494	0.8954	0.1921	0.3645	0.9983	0.3055	0.7605

G4	0.5855	0.2377	0.0971	0.0282	0.9090	0.0518	0.4206	0.7768	0.6902	0.5935
G5	0.5401	0.8144	0.8121	0.6938	0.6359	0.1169	0.9943	0.3467	0.5378	0.5618

Table 2: Initial expression values for 10 Data sets

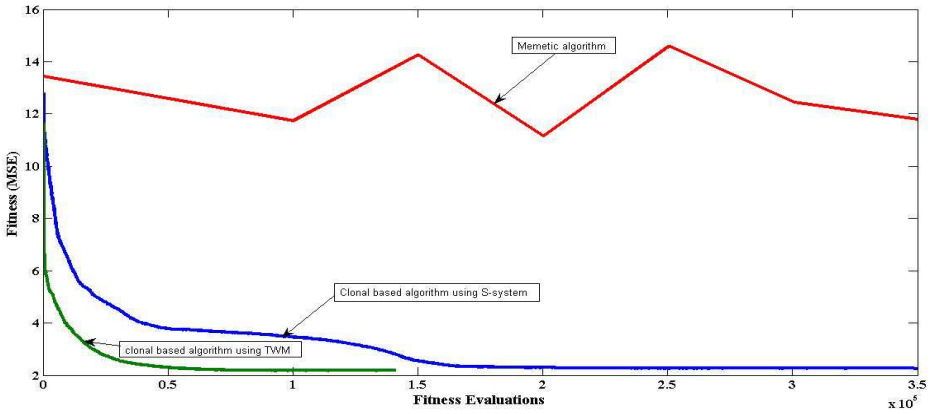

Fig. 4: Comparison of average errors (MSE) obtained for memetic algorithm, Clonal based algorithm using S-system and Clonal based algorithm using TWM model.

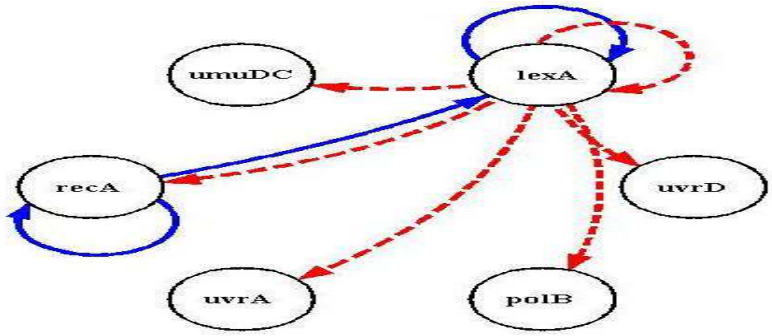

Fig. 5: SOS DNA repair system of E.coli.

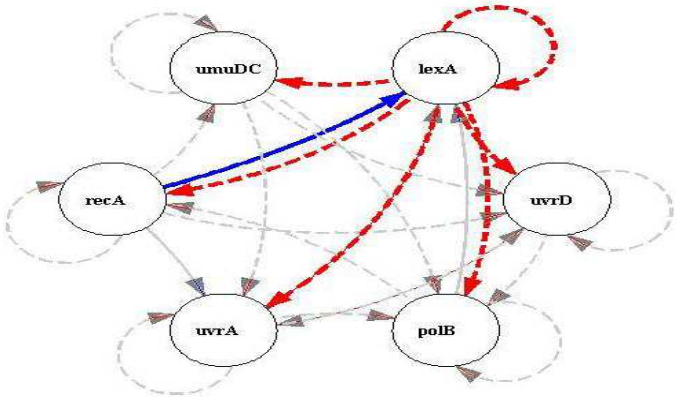

Fig 6: SOS DNA repair system of E.coli. Identified by the Clonal selection based approach using *Two-Weight-Matrix (TWM)* model

uvrD	uvrD-|uvrD[3,29,33,36,37], uvrD -> polB [3,34,35,36], uvrD-|uvrA[3].
LexA	LexA -| LexA[3,27,29,31,33,34,35,37], LexA -|uvrD[27,29,31,33,34,35,37], LexA -|uvrA[29,31, 33, 34, 35, 36, 37], LexA -|recA[3,27, 29,31,32,34,36,37], LexA -| polB[27,29,31,33,34,36], LexA -| umuDc[3,27, 29,31,32,33,34,36].
umuDc	umuDc-|umuDc[33,36], umuDc-|polB[34, 36], umuDc->uvrA[36].
recA	recA->uvrA[34,36], recA-|umuDc[29,35,36,37], recA -| recA[27, 29, 33, 36, 37], recA -> LexA[27, 29, 32, 33, 34, 36, 37], recA-|uvrD[35].
uvrA	uvrA-|uvrA [29, 33, 36, 37], uvrA -|lexA[35], uvrA -|polB[36], uvrA -|uvrD[27, 34, 35]
polB	polB-|polB[3,29,33,36], polB-|recA[27,36], polB->LexA[36].

Table 3: Relations identified by the proposed approach that are also already identified by previous researchers

Figure 6 shows the graphical representation of the gene regulatory network predicted by this proposed method. dotted lines show the inhibition and solid lines show the excitation relations. Since we are dealing with the real life microarray data sets, the given data is concealed noise. The accuracy of the prediction will be dependent on the degree of noise. Biological systems are too complex, hence even with the advanced technologies, biologist's knowledge about a biological system is limited. There is a possibility that the gene network identified by the biologist in Figure 3 may not contain all the relation of the DNA damage repair E.coli. System. A major application of the reconstruction of gene regulatory system is to identify such relations. The proposed method identified a set of relations among LexA to LexA, uvrD, uvrA, recA, polB, umuDc.

Table 3 shows the predicted relations by the proposed method. The proposed method identified a new relation, umuDc to uvrD, which may be a relation that is not identified by the researchers. From the table it can be predicted that there are some more relations which can be considered by the biologist when considering the system of Ecoli DNA repair.

[VI] CONCLUSION

Gene regulatory network represents the communication between genes to bring about a biological process. This helps to identify the importance of each gene in a particular biological action that happens in our body. It is evident that the researchers are yet to identify some more gene relations for many of the biological process. Researchers from the bioinformatics field are constantly trying to help biologists in the identification of the new gene relations and to confirm the existence of previously identified relations. As a part of this job, in this paper, we have introduced a better approach for the gene regulatory network reconstruction. The approach makes use of an evolutionary algorithm along with a novel model, called Two Weight Matrix model. The approach is implemented and compared with the standard memetic algorithm and S-systems, and the performances are found to be better than the existing approaches. The approach provides improved accuracy with faster convergence. The proposed approach is also implemented in a real life data set called SOS data of DNA repair system of E.coli. to demonstrate its applicability to real life data. The new proposal predicted most of the relations already identified by the researchers, and also predicted a new relation, umuDc to uvrD, which may be a relation that is not identified by the earlier researchers.

REFERENCES

1. Ando, S., Sakamoto, E., Iba, H., (2002), Modeling genetic network by hybrid gp, In:

Evolutionary Computation, 2002. CEC'02. Proceedings of the 2002 Congress on, IEEE. vol. 1, pp. 291–296.

2. Wang, H., Qian, L., Dougherty, E., (2006), Inference of gene regulatory networks using genetic programming and kalman filter, In: Genomic Signal Processing and Statistics, 2006. GENSIPS' 06. IEEE International Workshop on, IEEE. pp. 27–28.
3. Noman, N., Iba, H., et al., (2005), Reverse engineering genetic networks using evolutionary computation, GENOME INFORMATICS SERIES. 16(2), 205–214.
4. Savageau, M.A., years of s-systems. Canonical Nonlinear Modeling. S-systems Approach to Understand Complexity pp. 20:1–44.
5. Chowdhury, A.R., Chetty, M., (2011), An improved method to infer gene regulatory network using s-system, In: Evolutionary Computation (CEC), 2011 IEEE Congress on. pp. 1012–1019.
6. Chan, Z.S., Kasabov, N., Collins, L., (2006), A two-stage methodology for gene regulatory network extraction from time–course gene expression data, Expert Systems with Applications. 30(1), 59–63.
7. Sehgal, M., Gondal, I., Dooley, L., Coppel, R., (2006), Afegrn: Adaptive fuzzy evolutionary gene regulatory network re-construction framework, In: Fuzzy Systems, 2006 IEEE International Conference on. pp. 1737–1741.
8. Ram, R., Chetty, M., (2007), A guided genetic algorithm for learning gene regulatory networks, In: Evolutionary Computation, CEC 2007, IEEE Congress on. pp. 3862–3869.
9. Daisuke, T., Horton, P., (2006), Inference of scale-free networks from gene expression time series, Journal of bioinformatics and computational biology. 4(02), 503–514.
10. Ho, S.Y., Hsieh, C.H., Yu, F.C., Huang, H.L., (2007), An intelligent two-stage evolutionary algorithm for dynamic pathway identification from gene expression profiles, Computational Biology and Bioinformatics, IEEE/ACM Transactions on. 4(4), 648–704.
11. Shu, L.S., Huang, H.L., Ho, S.J., Ho, S.Y., (2011), Establishing large-scale gene regulatory networks using a gene-knowledge-embedded evolutionary computation method. In: Computer Science and Automation Engineering (CSAE), 2011 IEEE International Conference on. vol. 4, pp. 272–276.
12. Maraziotis, I., Dragomir, A., Thanos, D., (2010), Gene regulatory networks modelling using a dynamic evolutionary hybrid, BMC bioinformatics. 11(1), 140.
13. Mitra, S., Das, R., Banka, H., Mukhopadhyay, S., (2009), Gene interaction–an evolutionary biclustering approach, Information Fusion. 10(3), 242–249.
14. Schroder, A., Wrzodek, C.,Wollnik, J., Drager, A.,Wanke, D., Berendzent, K., Zell, A., (2011), Inferring transcriptional regulators for sets of co-expressed genes by multi-objective evolutionary optimization, In: Evolutionary Computation (CEC), 2011 IEEE Congress on. pp. 2285–2292.
15. Jereesh, A. S, Govindan, V. K., (2013), Gene regulatory network modelling using cuckoo search and s-system, International Journal of Advanced Research in Computer Science and Software Engineering. 3(9):1231–1237.
16. Jereesh, A. S, Govindan, V. K., (2013), A clonal based algorithm for the reconstruction of genetic network using S-system, International Journal of Research in Engineering and Technology (IJRET). 2:44–50.
17. Spieth, C., Streichert, F., Speer, N., Zell, A., (2004), A memetic inference method for gene regulatory networks based on s-systems, In: Evolutionary Computation, CEC2004. Congress on, IEEE. vol. 1, pp. 152–157.

18. de Castro, L.N., Von Zuben, F.J., (1999), Artificial immune systems: Part i–basic theory and applications, Universidade Estadual de Campinas, Dezembro de, Tech. Rep.
19. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., et al., (2002), The adaptive immune system.
20. Burnet, S.F.M., et al., (1959), The clonal selection theory of acquired immunity, Vanderbilt University Press Nashville.
21. Burnet, F.M., (1976), A modification of jerne's theory of antibody production using the concept of clonal selection, CA: a cancer journal for clinicians. 26(2), 119–121.
22. De Castro, L.N., Von Zuben, F.J., (2002), Learning and optimization using the clonal selection principle, Evolutionary Computation, IEEE Transactions on. 6(3), 239–251.
23. Tominaga, D., Koga, N., Okamoto, M., (2000), Efficient numerical optimization algorithm based on genetic algorithm for inverse problem, In: Proceedings of Genetic and Evolutionary Computation Conference. pp. 251–258.
24. Kikuchi, S., Tominaga, D., Arita, M., Takahashi, K., Tomita, M., (2003), Dynamic modeling of genetic networks using genetic algorithm and s-system, Bioinformatics. 19(5), 643–650.
25. Perrin, B.E., Ralaivola, L., Mazurie, A., Bottani, S., Mallet, J., dAlche Buc, F., (2003), Gene networks inference using dynamic bayesian networks, Bioinformatics. 19(suppl 2), ii138–ii148.
26. Kimura, S., Ide, K., Kashihara, A., Kano, M., Hatakeyama, M., Masui, R., Nakagawa, N., Yokoyama, S., Kuramitsu, S., Konagaya, A., (2005), Inference of s-system models of genetic networks using a cooperative coevolutionary algorithm, Bioinformatics. 21(7), 1154–1163.
27. Kimura, S., Sonoda, K., Yamane, S., Maeda, H., Matsumura, K., Hatakeyama, M., (2008), Function approximation approach to the inference of reduced ngnet models of genetic networks, BMC bioinformatics. 9(1), 23.
28. Noman, N., Iba, H., (2007), Inferring gene regulatory networks using differential evolution with local search heuristics, IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB). 4(4), 634–647.
29. Kabir, M., Noman, N., Iba, H., (2010), Reverse engineering gene regulatory network from microarray data using linear time-variant model, BMC bioinformatics. 11(Suppl 1), S56.
30. Tominaga, D., Koga, N., Okamoto, M., (1999), Efficient numeral optimization technique based on genetic algorithm for inverse problem, In: Proceedings of German Conference on Bioinformatics. pp. 127–140.
31. Sutton, M.D., Smith, B.T., Godoy, V.G., Walker, G.C., (2000), The sos response: recent insights into umudc-dependent mutagenesis and dna damage tolerance, Annual review of genetics. 34(1), 479–497.
32. Bansal, M., Della Gatta, G., Di Bernardo, D., (2006), Inference of gene regulatory networks and compound mode of action from time course gene expression profiles, Bioinformatics. 22(7), 815– 822.
33. Cho, D.Y., Cho, K.H., Zhang, B.T., (2006), Identification of biochemical networks by s-tree based genetic programming, Bioinformatics. 22(13), 1631–1640.
34. Kimura, S., Nakayama, S., Hatakeyama, M., (2009), Genetic network inference as a series of discrimination tasks, Bioinformatics. 25(7), 918–925.
35. Hsiao, Y.T., Lee, W.P., (2012), Inferring robust gene networks from expression data by a sensitivity based incremental evolution method, BMC bioinformatics. 13(Suppl 7), S8.

36. Huang, H.L., Chen, K.W., Ho, S.J., Ho, S.Y. (2008), Inferring s-system models of genetic networks from a time-series real data set of gene expression profiles, In: Evolutionary Computation, 2008. CEC 2008. IEEE World Congress on Computational Intelligence. pp. 2788–2793.
37. Mondal, B.S., Sarkar, A.K., Hasan, M.M., Noman, N., (2010), Reconstruction of gene regulatory networks using differential evolution. In: Computer and Information Technology (ICCIT), 13th International Conference on. pp. 440–445.