UDC 519.716.39: 519.6: 57.017

V.I. Lytvynenko, A.A. Fefelov,M.A. Taif, I.A. Lur`ie

A MULTI-OBJECTIVE IMMUNE APPROACH TO RECONSTRUCT GENE REGULATORY NETWORK USING ALGORITHM CLONAL SELECTION

Abstract: *The inference of gene regulatory networks is one of the main challenges in systems biology. In this paper we address the problem of finding gene regulatory networks from experimental DNA microarray data. We suggest to use a multi objective clonal selection algorithm to identify the parameters of a non linear system given by the observed data. Not only the actual parameters of the examined system are unknown, also the connectivity of the components is a priori not known. However, this number is crucial for the inference process. Consequently we propose a method based on algorithms of artificial immune system which uses the connectivity as an optimization objective in addition to the data dissimilarity (relative standard error RSE) between experimental and simulated data.*

Keywords: *gene regulatory networkst; multi objective optimizacion; clonal selection algorithm; SOS DNA repair network.*

1. INTRODUCTION

Gene Regulatory Networks (GRNs) are the functioning circuitry in living organisms at the gene level. It is regarded as an abstract mapping of the more complicated biochemical network which includes other components such as proteins, metabolites, etc. The purpose of GRN is to represent the regulation rules underlying the gene expression. Understanding GRNs can provide new ideas for treating complex diseases and breakthroughs for designing new drugs [1].Gene regulatory network reconstruction is currently a topic under heavy research in the computational biology field. The study of GRN is made much easier with the recent introduction of *microarray* technology. Using this method, expression levels of thousands of genes can be measured simultaneously, as they change over time and are affected by different stimuli. Thereby, it is possible to obtain a global view of the dynamic interaction among genes. But it is a great challenging problem to discover these networks of interacting genes that generate the fluctuations in the gene expression levels [1]. Gene regulatory networks are distributed system of genes in various combinations. The interaction of genes controls the biological structure and functions of proteins. Properties of a gene undergo changes when it comes in the presence of other genes. Identifying a gene network is a complex nonlinear problem. Before the invention of microarray

© Lytvynenko V. I., Fefelov A.A., Taif M. A., Lur`ie I. A., 2015

technology, parallel processing of the genes was a complex and difficult problem for the biologists. The introduction of microarrays transformed the problem of gene profiling to a simple and efficient one. For example, microarray representation may be used for clustering genes and reconstruction of gene network making use of gene expression time series. Reverse engineering approach for the identification of gene networks is a well-accepted approach in the literature. In this, gene expression profiles identified by the microarray are used to predict the gene regulatory network [2] . Inference of GRNs based on microarray data is referred to as *reverse engineering* [3], as the microarray expression levels are the outcome of gene regulation. Mathematically, *reverse engineering* [RE] is a traditional inverse problem.

Reverse Engineering (RE) can be considered as a process from which is possible inferring structural and dynamics feature of a given system from external observations and relevant knowledge. Thanks to this feature, today RE techniques play a central role in systems biology [4-5], since it is not only important a knowledge of genes and proteins, but also to understand their structures and dynamics [6].

The solution to the problem is, however, not tri vial, as it is complicated by the enormously large scale of the unknowns in a rather small sample size. In addition, the inherent experimental defects, noisy readings, and many other factors play a role. These complexities call for heavy involvement of a powerful mathematical modeling together with reliable inference, which play an increasingly important role in this research [1].

Many types of models have been already proposed for reconstruction of gene regulatory networks in biological systems, including Boolean networks [7], linear weighting networks [8], differential equations [9], Bayesian Networks [10], S-system [11], Fuzzy Set [12] and Artificial & Recurrent Neural Network [13] etc. Researchers have proposed a number of evolutionary algorithms for the construction of gene regulatory networks. In work [14] 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. In work [15] has proposed a ***Genetic Program*** proposed which makes use of ***Kalman filter*** for estimating the parameters 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. In work [16] 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 [17] 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. 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 proposed in work [18]. This approach is based on the ***fuzzy clustering using EA*** and **Spearman correlation.** In [19], authors considered biological network as a scale free network and used advancement of GA called ***Distributed genetic algorithm*** to optimize S-system parameters. The issue here is that the knowledge about such properties are often not available. Gene Regulatory Network Modeling using **Cuckoo Search** and **S-system** [20] 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 [21]. In work [22] 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) [22]. 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 [17]. 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).

Currently, most of the existing algorithms developed for the reconstruction of the GRN are single-objective [23].

However, previous research on single objective GRN showed that on single objective can generate similar results of experimental data, but they may not be numerical or structural similarity with real network [24-26].

This may occur because the optimization process is caught in local optima. Stochastic multi-objective approaches preserve the diversity of solutions in a population and present them as a Pareto front. Thus they are able to find multiple optima hopefully including the global optimum [27]. The multi-objective optimization approach is likely to be more suitable for genetic regulatory modeling and its associated parameter estimation based on following three reasons [28]:

a) Multiple data types (continuous, discrete, and/or categorical) are very problematic for the design of a single objective function;

b) Individual data sets usually are from different sources and may be inconsistent;

c) Tradeoffs between solutions may reveal the magnitude of discrepancies.

On the basis of the above reasons, research on multi-objective algorithms in modeling gene regulatory network is relatively new but rapidly growing area of research. There are few attempts to use more objective approach to GRN have been described in [27-28]. Further more, previously work on this topic showed that, due to the multi-modal character of the solution space, several sets of parameter exist, which fit the data satisfactorily. Thus, standard optimization techniques are easily caught in local optima, i.e. finding a solution with a good RSE but with no structural resemblance with the true system. This is known to be a major problem in the inference process [23-28]. Because multi-objective algorithms preserve the diversity of the solution within a population by maintaining the Pareto-front and are therefore able to find multiple optima hopefully including the global optimum. All work on the application of multi-objective algorithms for reconstruction of GRN apply genetic algorithms. Researches have shown the advantage of artificial immune system algorithms to genetic algorithms, since the first combine local and global search. For that reason, this paper is devoted to application of clonal selection algorithm to solve the problem of

reconstruction GRN using multi-objective optimization based on s-systems.

2. PROPOSED METHOD

2.1. The Genetic Network Model and S-System

While there are a series of attempts already been conducted by different researchers, received all the solutions are still not satisfactory, that in relation to the time required and the achieved degree of accuracy. Therefore there is need for further research on this topic to reach satisfactory solutions with improved performance. GRNs network describe bimolecular interactions that are inherently non-linear and can be expressed by a common system of differential equations. Biochemical systems like GRNs are generally modeled by systems of ordinary differential equations (ODEs). However, nonlinear differential equation models, such as S-system, can model much more complicated GRN behavior successfully. In general, modeling GRNs may be considered as a nonlinear identification of dynamics problem. If there are N genes of interest; define xi as the state (such as the gene expression level) of the i-th gene, then the dynamics/interactions of the GRN may be modeled as

$$\frac{dx_i(t)}{dt} = f_i[x_1(t),\ldots,x_N(t)] \quad (1)$$

for i = 1,...,N, where N this number of genes, x_i – gene expression level, but f_i –a function that describes the influence of genes on a gene i. . For example, if j-th gene activates i –th gen, the value $f_i$increases with x_j and conversely, if j-th gene inhibits i-th gene.

The proposed method used by S- system, which is widely recognized as a model for the reconstruction of gene regulatory networks .

One of the ways the mathematical description of a genetic network is the S-system [17], which is a system of differential equations of the form:

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^{n} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n} X_j^{h_{ij}} \quad (2)$$

where n - number of state variables that characterize the investigated object or the number of reagents X_i. $\alpha_i$and $\beta_i$are the positive rate constants for increasing and decreasing respectively. $g_{i,j}$ and $h_{i,j}$ are the exponential parameters that are also called as kinetic orders. If $g_{i,j} > 0$, gene j will excite the expression level of gene i. On the other hand, gene j will inhibit

the expression level of gene i if $g_{i,j} < 0$. $h_{i,j}$ have the inverse effects on controlling gene expressions compared to gi,j. S-system model is characterized by power-law functions and it has the rich configuration capability of capturing various dynamics in many complex biochemical systems. As the S-system model has been proven to be successful in modeling GRNs.

2.2 Multi-objective criterion [29]

For examining the connectivity and the RSE in parallel, we used a multi-objective EA, which optimizes the parameters of g, h, α_i and β_i in respect to the following two optimization objectives:

a.) For evaluating the RSE fitness of the individuals we used the following

equation for calculation of the fitness values:

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

where N is the total number of genes in the system, T is the number of sampling points taken from the experimental time series $X_{i,cal;t}$ - the level of expression of the gene X_i-th gene at time t is calculated numerically by solving a system of differential equations (1) for the intended set of parameters and $X_{i;\ exp;\ t}$ is experimentally observed gene expression level of X_i in time $\boldsymbol{t}$.

The problem is to minimize the fitness value f_1.

b) The second optimization objective is to minimize the connectivity of the system, as biologically the gene regulatory network is known to be sparse. The connectivity is defined in two different ways: first, the maximum connectivity of the genes, i.e. the total number of interactions of the system:

$$f_2^a = \sum_{i=1}^{N} \left(\left| sign(\alpha_i) \right| + \left| sign(\beta_i) \right| \right) + \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\left| sign(g_{i,j}) \right| + \left| sign(h_{i,j}) \right| \right) \quad (4)$$

And secondly, the median average connectivity of all genes, i.e. the median

average number of interactions of each gene:

 ISSN 1562-9945

$$f_2^b = median\left(\frac{f_2^a}{N}\right) \quad (5)$$

The proposed method is based on an algorithm of artificial immune network (AIS). The structure of the genome (the antibody Ab) is represented as a set of values of the optimization parameter $\alpha_i, \beta_i, g_{ij}, h_{ij}$.

2.3. Clonal Selection Algorithm

When solving the optimization problem, the goal is to find the optimal values (minimum or maximum) of some criterion, $y = f(f_1, f_2^b)$, $x_i \in X$, $i = \overline{1,l}$ where X - the feasible set of tasks. In general, we consider problems of multi-objective optimization:

$$y = (f_1, f_2^b) \to min \ , \quad (6)$$

where $y_j = f_j(f_1, f_2^a, f_2^b)$, $j = \overline{1,n}$ - number of objective for the task. Depending on the conditions of the problem finding possible global or local optima.

In optimization problems, the generalized form of antibodies is a vector of arguments $Ab = (x_1, x_2, ..., x_l)$, and as antigens used optimality criteria y_j, expressed as functions $Ag = f(x_1, x_2, ..., x_l)$. Affinity values g_j calculated on the basis of criteria values y_j reflected in the set of nonnegative numbers such as:

$$f : X \to \Re, \quad F : \Re \to \Re^+ \quad (7)$$

Thus, there is some affinity function $g = F\big(f(x_1, x_2, ..., x_n)\big)$, that determines that determines the degree of conformity of individuals to each other. In such problems, we can not to operate the notion of distance, so that the best value criteria is previously unknown, and, therefore, we do not know the maximum possible extent to which individuals. Thus, the control dynamics of AIS is performed by the relative affinity values or by rank individuals set. This approach is very close to the concept of suitability (fitness) used in evolutionary algorithms that have some earlier theory of artificial immune systems

Formally algorithm of clonal selection can be represented as:

$$CLONALG = (P^l, G^k, l, k, m_{Ab}, \delta, f, I, \tau, AG, AB, S, C, M, n, d), \quad (8)$$

where P^l is space of search (space of forms); G^k is space representation; l is the length of vector of attributes (dimension of space of search); k is the length of antibody receptor; m_{Ab} is dimension of population of antibodies; δ is the expression function; f is the affinity function; I is the function of initialization of the initial population of antibodies; τ is the condition of completion of algorithm work; AG is the subset of antigenes; AB is population of antibodies; S is the operator of selection; C is the operator of cloning; M is the mutation operator; n is the number of the best antibodies selected for cloning; d is the number of the worst antibodies subjected to substitution for new ones.

Consider the shape- space $\left(P^l\right)$ phenotypes and their space images as antibodies $\left(G^k\right)$ or genotype space .

Function:

$$\delta : P^l \to G^k \quad (9)$$

is a function of conversion options with P^l solutions to their internal image G^k in a population of individuals.

This function is another called function expression. It should be noted that in practice, the development AIS often impose similar transformation for reasons of convenience, the application of immune operators and calculate affinity individuals.

For example, a vector of real attributes of dimension l can be transformed into a bit string of length k, which enables the use of specific operators mutation and affinity calculations using different types of Hemming distance.

Therefore, the terms "*genotype*", "*phenotype*" and "*expression*" in this description are borrowed from relatives by their functional structure and evolutionary algorithms, although more suitable for use in the context of the evolution of chromosomes than the molecular structure of antibodies. It is assumed also that for every solution $p \in P^l$ there is one and only one of his images $\delta(p) \in G^k$. Thus in general the opposite assertion is not correct.

Using a generalized mapping [29], you can type affinity function f:

 ISSN 1562-9945

$$f : P^l \times P^l \to \Re^+ . \quad (10)$$

This problem is to maximize the function of affinity. Taking the initial population size of antibodies (m_{Ab}), you can enter the initialization function as:

$$I : G^k \times m_{Ab} \to AB\left(G^k\right). \quad (11)$$

Often, the initialization is carried out randomly with uniform distribution.

Let the Q - unary operator stochastic transformation on the set G^k, which uses administering set K_Q to generate control parameters that determine the way to convert the current step of the procedure. For example, if mutations bit string, bit mask can be used as a control parameter, which determine the position of individual numbers of the individual bits that undergo mutations. Thus, the functional entry operator Q can be represented as follows:

$$Q : G^k \times K_Q \to G^k. \quad (12)$$

The optimum solution of $Ab_{opt} \in G^k$ concerning the operator of Q and antigen $Ag \in AG, \quad AG \subset G^k$ is called an individual, whose affinity can't be increased at further influence of the operator of transformation Q, ie

$$\forall k \in K_G : f\left(Q(Ab_{opt}, k), Ag\right) \le f\left(Ab_{opt}, Ag\right). \quad (13)$$

The condition of shutdown (τ) is executed when the population of antibodies fully recognizes the population of antigens, ie

$$\forall Ag \in AG : \exists Ab \in G^k \mid Ab = Ab_{opt}. \quad (14)$$

Operator selection S forms a subset of G_S individuals whose affinity is better in this generation. Thus, S together with the management set K_S represents the function:

$$S : G^k \times K_S \to \{0, 1\}, \quad (15)$$

set, which is formed by the selection:

$$G_S = \left\{Ab \in G^k \mid S(Ab, k_S) = 1\right\}, \quad \left|G_S\right| = n. \quad (16)$$

Similarly, the selection is made of individuals in a population of cells of memory. Operator cloning C increases of elements of the set G_S in population and with the management set K_C can be written as:

$$C : G_S \times K_C \to G_S . \tag{17}$$

The operator of a mutation of M with operating set K_M:

$$M : G^k \times K_M \to G^k . \tag{18}$$

Metadynamic system expressed as a function of substitution worst antibody population:

$$R : G^k \times d \to AB_d\left(G^k\right). \tag{19}$$

Worse antibodies previously selected by the operator selection.

Model process of transformation states populations clonal antibodies using procedures:

$$AB_t \xrightarrow{Selection(S)} G_S \xrightarrow{Cloning(C)} G_C \xrightarrow{Mutation(M)}$$
$$G_M \xrightarrow{\text{Repeat mutation}(S)} G_S \xrightarrow{\text{Replacement}(d)} AB_{t+1},$$

where t – number generation; AB – population of antibodies (detectors); G_S – best subset of selected antibodies; G_C – a subset of the clones; G_M – a subset of clones after mutation.

Let us show the generalized stepwise description of the algorithm.

1. *Initialization.* Creation (usually by random generation) of the initial population of antibodies AB.

2. *Determination of affinity.* For every antibody AB_j, $AB_j \in AB$ determine its affinity relative to every antigen Ag_i, $Ag_i \in AG$. Write the result into the matrix of affinities $D : D = \left[\left|AG\right| \times m_{Ab}\right]$, and $d_{ij} = f\left(Ab_j, Ag_i\right)$, $d_{ij} \in D$.

3.*Clonal selection and propagation.* Select from population n of each the best antibodies for every row of the matrix D and place them into separate population of clones $AB_C, \left|AB_C\right| = n \cdot \left|AG\right|$. It is necessary to generate clones of elements of the population AB_C proportionally to their affinity, i.e., the greater it is, the greater number of clones is generated and vice versa.

 ISSN 1562-9945

4. *Affinity maturation.* Subject to mutation all the clones of population AB_C with probability inversely proportional to affinities, i.e., probability of mutation is the greater, the lower is its affinity. Determine new affinity of every antibody AB_j, $AB_j \in AB_C$ similar to item 2 and obtain the matrix of affinities D_C. Select n antibodies from the population AB_C, for which the corresponding vector-column of the matrix D_C gives the best generalized result of affinity, and transfer them into population of cells of memory M_R.

5. *Metadynamics.* Substitute the worst d antibodies of the population AB by new random individuals.

6.Substitute n antibodies of the population AB by cells of memory from M_R and pass to item 2 until the stoppage criterion is reached.

The operator of cloning used to create a set of copies of the best individuals in the population . This operator makes it possible to increase the study intensity of solutions area space. Cloning (generation of identical copies) is the selection of an individuals quantity, while the quantity of copies in proportion to their affinity: the higher the affinity, the larger clone (offspring quantity). In optimization tasks the n antibodies with highest affinity are shown. The number of clones is calculated by the formula:

$$N_c = \sum_{i=1}^{N} round(\beta \cdot N) , \qquad (20)$$

where N_c is the total number of clones created for each of the antigens-multiplying factor, N - is the total number of antibodies, $round(\cdot)$ - rounding operator of argument to an integer.

Each antibody is seen in the local scale and does not have any advantages when cloning to other antibody. Antigenic affinity (matching the value of objective function) further is used to determine the hyper mutation level for each clone of the antibody.

After cloning operation, the clones are subject to a hypermutation process inversely proportional to their affinity; the higher the affinity the smaller the mutation rate. Hence, the mutation rate of a clone is inversely proportional to the fitness of its parent. The antibodies in each subpopulation which consists of the parent and its clones maturated by

hypermutation operation are then evaluated in the affinity function, and the best antibody of each subpopulation becomes memory cell and is allowed to survive. The antibodies with d lowest affinities are replaced by the new antibodies generated randomly to maintain the diversity of antibody population so that the new areas of the search space can be potentially explored. The next generation starts with a new antibody population produced as described above. These processes are repeat d until a termination criterion is attained or a predetermined generation number is reached [30-32].

Principle 1: *The proliferation rate of each immune cell is proportional to its affinity with the selective antigen (higher the relative affinity, themoreprogeny)* [33].

Principle 2: *The mutation suffered by each immune cell during reproduction is inversely proportional to the affinity of the cell receptor with the antigen (higher the relative affinity, the lower the mutation)* [33].

Each candidate solution (an attribute string in a given shape space) has an independent mutation rate in proportion to its affinity with the optima solutions. Thus, candidates in higher peaks of the affinity land will be subject to smaller mutation rates while candidates located far from optima solutions will suffer larger mutation rates. One problem with this approach is that, usually, nothing is known a priori about the optima solutions of a problem In this case, one can evaluate the relative affinity of each candidate solution by scaling (normalizing) their affinities. The inverse of an exponential function can be used to establish a relationship between the hypermutation rate $\alpha(\cdot)$and the normalized affinity D^* [30-32]. In some cases it might be interesting to re-scale a to an interval such as [0..0.1].

$$\alpha\left(D^*\right) = \exp\left(-\rho D^*\right), \tag{21}$$

where ρ is a parameter that controls the smoothness of the inverse exponential, and D^* is the normalized affinity, that can be determined by $D^* = D / D_{\max}$.

3. EXPERIMENTS AND RESULTS

This 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. It was taken from www.weizmann.ac.il/mcb/UriAlon/Papers/SOSData/. In this system, there are six major genes that control the process of repair. Those genes are uvrD, LexA, umuD, recA, uvrA and polB. The data set obtained was normalized before using it for the prediction of relations in the gene network. In the SOS DNA repair system of *Ecoli*. LexA is a suppressor of all the other genes. Whenever DNA damage occurs, the concentration of LexA drops. This activates all other genes and starts repairing. After repairing, LexA gets back to original position and all the other genes are suppressed. Thus, the system attains a stable state.

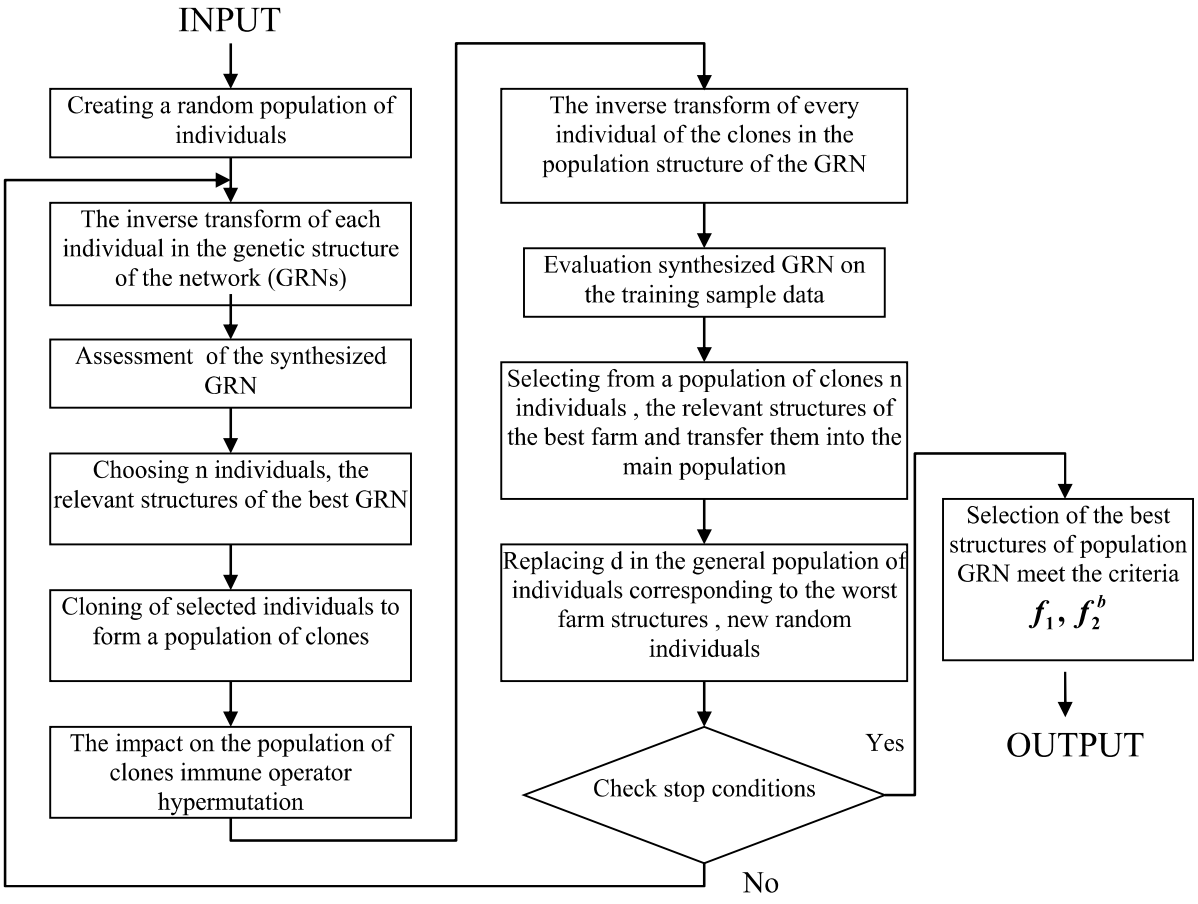

Fig. 1. Workflow of multi-objective clonal algorithm optimization for the identification of a gene regulatory network

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.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].

Table 1 shows the results of reconstruction of the gene regulatory network using as an objective function f_1

Table 1.

The results of the application of clonal selection algorithm for single optimization

	ubrD	lexA	umuD	recA	uvrA	polB
ubrD	0	0	1	1	0	0
lexA	1	1, -1	0	1, -1	1, -1	-1
umuD	1, -1	0	1	1	1	
recA	-1	1, -1	-1	0	1	1
uvrA	1	1	0	0	-1	1
polB	0	0	0	0	0	-1

In Table 2 shows the results of the reconstruction of gene networks using multi-objective function $y = (f_1, f_2^b) \to min$

Table 2.

The results of the reconstruction of gene networks using clonal selection algorithm with the application of multi-objective optimization

	ubrD	lexA	umuD	recA	uvrA	polB
ubrD	-1	-1	-1	0	0	1
lexA	-1	-1	0	-1	1	1
umuD	0	-1	-1	-1	1	-1
recA	0	-1	-1	1	1, -1	-1
uvrA	1	-1	1	-1	-1	-1
polB	1, -1	0	-1	-1	1, -1	-1

Using clonal selection algorithm with the application of multi-objective optimization predicted the major biological relations such as lexA to lexA, umuDc, recA, and uvrA. It also identified major positive

 ISSN 1562-9945

relation from recA to lexA, recA to recA and lexA to lexA. The proposed clonal selection algorithm identified two biological relations (recA to recA, lexA to umuDc) that are not identified by the clonal selection algorithm based S-system approach. The existing clonal selection algorithm based S-system approach predicted six biologically identified relations and the proposed approach identified seven biologically identified relations. An advantage of the proposed approach is that the number of uncertain relations identified is much lower than that of clonal selection algorithm based S-system approach. Uncertain relations are the relations that are not sure whether they exist. For the clonal selection algorithm based S-system approach, the numbers of uncertain relations are 22 and for the proposed clonal selection algorithm are 12. This clearly indicates that the proposed approach can successfully use for the real life data set, and it is efficient in predicting relation than the clonal selection algorithm based S-system.

4. CONCLUSION

In this work, we presents an immune multi-objective approach to inference network of S-system, using the clonal selection algorithm based S-system. Due to the proposed multi-objective optimization model, there is no need to preset any parameter value before the inference MOEA is run. Thus, our method could be generally applicable to various kind of GRN model. Future work will be focused on inference of GRN based on noisy data, and how to generalize the proposed method to infer big-scale GRNs.

REFERENCES

1.Kabir, M. Noman, N. Iba, H., Reverse engineering gene regulatory network from microarray data using linear time-variant model, BMC bioinformatics 11 (Suppl 1) (2010) S56.

2. Jereesh A S, Govindan V K An effective reverse engineering methodology for the reconstruction of gene regulatory network using gene expression profile," International journal of Universal Pharmacy and Biosciences, vol. 2, no. 6, pp. 32-42, 2013

3. D'haeseleer P, Liang S, Somogyi R: Genetic network inference: from co-expression clustering to reverse engineering. Bioinformatics 2000, 16(8):707–726. 10.1093/bioinformatics/16.8.707

4. Csete M., Doyle J.: "Reverse Engineering of Biological Complexity," Science, vol. 295, pp. 1664–1669, 2002.

5.Lee D.: ”Component-Based Software Architecture for Biosystem Reverse Engineering,” Biotechnology and Bioprocess Engineering, vol. 10, pp. 400–407, 2005.
6. M. Pavone, P system Reverse Engineering for Gene Regulatory Networks in S-system Model, Unifying Themes in Complex Systems - Volume VIII, proceedings of the Eighth International Conference on Complex Systems (ICCS), New England Complex Systems Institute, pp. 492-503, Quincy, MA, USA, June 26 - July 1, 2011.
7. T. Akutsu, S. Miyano, and S. Kuhara, "Identification of Genetic Networks from a Small Number of Gene Expression Patterns under the Boolean Network Model," Pacific Symposium on Biocomputing 4, pp. 17-28, 1999.
8. D. C. Weaver, C. T. Workman, and G. D. Stormo, "Modeling Regulatory Networks with Weight Matrices," Pacific Symposium on Biocomputing 4, pp.123, 1999.
9. T. Akutsu, S. Miyano, and S. Kuhara, "Algorithms for Inferring Qualitative Models of Biological Networks," Pacific Symposium on Biocomputing 5, pp. 293-304, 2000.
10. K. Murphy and S. Mian, "Modelling Gene Expression Data using Dynamic Bayesian Networks," Computer Science Division, University of California, Berkeley 1999.
11. H. Wang, L Quin and E. Dougherty, “Inference of Gne Regulatory Network using S- System: A unified Approach”, Proceeding of 2007 IEEE symposiam CIBCB, pp. 82-89, 2007.
12. P. Du, J. Gong, E.S. Wurtele, and J.A. Dickerson, “Modeling gene expression networks using fuzzy logic,” IEEE Transactions on Systems, Man and Cybernetics, vol. 35, pp. 1351-1359. 2005.
13. J. Vohradsky, “Neural network model of gene expression”, The FASEB Journal, Vol. 15, pp.846-854, 2001.
14. 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.
15. 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.

16. Noman, N., Iba, H., et al., (2005), Reverse engineering genetic networks using evolutionary computation, GENOME INFORMATICS SERIES. 16(2), 205–214.
17. Savageau, M.A., years of s-systems. Canonical Nonlinear Modeling. S-systems Approach to Understand Complexity pp. 20:1–44.
18. Sehgal, M., Gondal, I., Dooley, L., Coppel, R., (2006), Afegrn: Adaptive fuzzy evolutionary gene regulatory network reconstruction framework, In: Fuzzy Systems, 2006 IEEE International Conference on. pp. 1737–1741.
19. 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.
20. 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.
21. 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.
22. 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.
23.Cai, X.: A multi-objective gp-pso hybrid algorithm for gene regulatroy network modeling. Ph.D. thesis, Kansas State University, Manhatten, Kansas (2009)
24. J. Herz, “Statistical issues in reverse engineering of genetic networks,” in Proceedings of the Pacific Symposium on Biocomputing, 1998.
25. C. Pridgeon and D. Corne, “Genetic network reverse engineering and network size: can we identify large GRNs?” In Proceedings of the Computational Intelligence in Bioinformatics and Computaitonal Biology (CIBCB’2004), pp. 32-36, 2004.
26. I. Ono, R. Yoshiaki Seike, N. Ono and M. Matsui, “An evolutionary algorithm taking account of mutual interactions among substances for

inference of genetic networks," In proceedings of the IEEE Congress on Evolutionary Computation (CEC' 2004), pp. 2060-2067, 2004.
27. C. Spieth, F. Streichert, N. Speer and A. Zell, "Multi-objective model optimization for inferring gene regulatory networks," Lecture Notes in Evolutionary Multi-Criterion Optimization, vol. 3410, pp. 607-620, 2005.
28.P. Koduru, Z. Dong, S. Das, S. M. Wlech, J. Roe and E. Charibit "Multi-Objective Evolutionary-Simplex Hybrid Approach for the Optimization of Differential Equation Models of Gene Networks", IEEE Transactions on Evolutionary Computation, accepted.
29. C. Spieth, F. Streichert, N. Speer, and A. Zell Multi-Objective Model Optimization for Inferring Gene Regulatory Networks Conference on Evolutionary Multi-Criterion Optimization (EMO 2005), Guanajuato, Mexico LNCS 3410, pp. 607–620, 2005, Springer-Verlag
29-. Bidyuk P.I. , Lytvynenko V.I., Fefelov A.F. The formalization of methods for building artificial immune systems // Science Visti of the National Technical University of Ukraine "Kyiv Polytechnic Institute"/, 2007, № 1, 2007, -P. 29-41. (in Ukrainian)
30. De Castro L.N. Artificial Immune Systems: Part I – Basic Theory and Applications/ De Castro L.N., Von Zuben F.J. (Technical Report – RT DCA 01/99, FEEC/UNICAMP), Brazil, 1999. – 95 p.
31. 10. De Castro, L. N. and J. Timmis Artificial Immune Systems: A New Computational Intelligence Approach./. Springer, Heidelberg, 2002. – 357 p.
32. De Castro L.N. Artificial Immune Systems: Part II – A Survey of Applications./ De Castro L.N., Von Zuben F.J. // Technical Report – RT DCA 02/00, FEEC/UNICAMP, Brazil. 2000. - 64 p.
33 Brownlee J. Clonal Selection Theory & CLONALG - The Clonal Selection Classification Algorithm (CSCA) // Victoria, Australia: Centre for Intelligent Systems and Complex Processes (CISCP), Faculty of Information and Communication Technologies (ICT), Swinburne University of Technology; 2005 Jan; Technical Report ID: 2-01, -P. 43.