A New Artificial Genetic Regulatory Network Model and its Application in Two Dimensional Robot Control

Navid Rahbari Asr and Vahid Johari Majd

***Abstract*—A new artificial genetic regulatory network model is introduced and an application of it is presented through controlling a two dimensional robot. The model is a refined and generalized version of previously introduced models. In a case study the new model has been able to create proper behaviors needing less number of parameters compared with a previous one.**

***Index Terms*—Genetic regulatory networks, genetic algorithms, genes, proteins, robot control.**

I. Introduction

Genetic Regulatory Networks (GRNs) are the basic control structures of biological cells. They consist of genes, proteins and metabolites that interact with each other and control the cell's behavior. [1] Modeling these structures has great value from two different aspects. On one hand it can help researchers to understand the underlying interconnection of cell functions which can further help them to find solutions to some genetic engineering problems. On the other hand artificial systems modeling genetic regulatory networks can be helpful for engineers to invent GRN inspired tools for manipulating control systems intelligently. In this usage they might have a potential to be as handy as some of the previously developed nature inspired tools like Neural Networks, Genetic Algorithms and Fuzzy logic.

Thus far there have been two different trends in literature toward modeling GRNs. On one side there are attempts to model GRNs with previously known structures such as deterministic and probabilistic Boolean networks [2, 3, 4, 5] analogue networks, differential equations [6], discrete event systems, Petri nets [7] and Markov chains [8]. These models are mostly helpful in biological systems identification and reverse genetic engineering problems. On the other side there are the Artificial Life (ALife) researches which attempt to define artificial dynamical cells with artificial genome, artificial proteins and artificial chemistry to simulate the behaviors of biological cells in silico. [9]-[11]. These models have the capability of developing GRN inspired controllers and in the recent decade some works have been done in this end.

In the recent decade some interesting researches have been performed in the area of GRN inspired controllers [12]-[15]. In [13], [14] artificial cells are evolved to have a coupling relationship with their environment and to control some robot behaviors. The properties of the artificial cell models in [13], [14] are that the number of different proteins is fixed and equals to the number of different genes, there is no interaction between proteins and the cell receives a feedback from its controlling behavior. In [15] a range of artificial cells have been evolved to control a robot on a two dimensional plane. In the model of [15] there is a dynamic interaction between proteins and the number of possible different proteins is much larger than the number of different genes. However the model lacks feedback from its controlling environment. There are also some discontinuities in protein interactions. For example concentration level of interacting proteins does not have any effect on the shape of the resultant proteins.

In this paper by considering the advantages and disadvantages of the three models described briefly in the previous paragraph, we have developed a model which tries to capture the strong aspects of the three models. Special features of our model are:

1) Number of possible proteins is infinite. They are regularly generated and decayed in the cytoplasm.
2) Different proteins interact with each other and form new proteins.
3) The discontinuities of protein interactions of the model in [15] are solved.
4) A feedback passage is enabled for the cells.

The paper is organized as follows. Section2 describes the proposed artificial cell model. Section3 reports some experimental results done with the new model and the final section summarizes the experimental results and suggests future work.

II. Artificial Cells

The artificial cells we introduce here are composed of three main parts:

1) Genome
2) Cytoplasm
3) Proteins

Genome is the place where genes are located. Genes interact with proteins in the cytoplasm and affect cell behavior. Proteins are structures that are created by genes and interact with genes and with each other. The cell is located in the environment.

1) Genes

The genes on the genome can be of one of the five types listed below:

1) Environment genes: these genes determine which proteins will be present in the environment of the cell.

Manuscript received September 20, 2012; revised October 30, 2012

The authors are with the Department of Electrical Engineering, Tarbiat Modares University, Tehran, Iran (e-mail:navidrahbariasr@gmail.com. majd@modares.ac.ir).

DOI: 10.7763/IJIEE.2013.V3.358

2) Cell receptor genes: these genes define which portions of the environment proteins are permitted to enter the cytoplasm.
3) Behavioral genes: when these genes are activated they cause a cell behavior. The behavior can be a cell production, execution of a function, enabling an event etc. It is through behaviors that the cell connects with the desired control space and performs the control action.
4) Regulatory genes: when these genes are activated they change the rate of production of proteins associated with them.
5) Input genes: these genes are activated or repressed from the control space and when activated they release proteins to the cytoplasm. These genes were not included in the model of [15] and they are added to enable a feedback control input for the cell.

2) Proteins

Proteins are structures that interact with each other and with coding and regulatory regions of a gene. In [15] these proteins are defined to be subsets of the Mandelbrot set. However we have defined a protein of order n to be an arbitrary array of size n whose elements are between zero and one. With this definition, the number of all possible proteins of order n would be infinite.

3) Chemistry of the artificial cell

We cover this section in two parts. Protein-Protein interactions and Protein-Gene interactions. But first, two notions necessary to proceed with the section are introduced. Degree of similarity and level of concentration.

Definition 1.Degree of similarity between two proteins:

Degree of similarity between two proteins P_1 and P_2 of order n, denoted by $DS(P_1,P_2)$, is defined by relation (1). In this way when two proteins are exactly the same, their degree of similarity would be one and when one of them is composed of only ones and the other is composed of only zeros, their degree of similarity would be zero.

$$DS(P_1,P_2)=1-\frac{1}{n}\sum_{i=1}^{n}|P_1(i)-P_2(i)| \quad (1)$$

Definition 2.Concentration of a protein:

To every protein in the cell environment, a concentration level is assigned. In [15] this level is defined to be between zero and 200. Without loss of information and to normalize the model, we assign a number between zero and one to the concentration level. Zero means the protein does not exist in the environment and one means the protein is saturated. So a concentration level of 100 in the model of [15] would correspond to 0.5 in our model. The concentration level of protein P_1 is denoted by C_{P_1}.

4) Protein-protein interactions

Proteins existing inside or outside the cytoplasm of a biological cell interact with each other to form new proteins. In artificial cells defined here, proteins inside or outside the cytoplasm are merged together to form a new protein as well. This composition of proteins is declared in the following definition.

Definition 3.Composition of proteins:

If proteins P_1 , P_2 , … , P_n with concentration levels C_{P_1} , C_{P_2} , … , C_{P_n} exist in a common environment (inside the cytoplasm of a cell or outside the cytoplasm) they are composed together to form a new protein P whose shape and concentration level is computed by the rules (2) and (3) respectively. In rule (2), $P(i)$ is the i th element of protein P ..

$$P(i)=\begin{cases}\dfrac{C_{P_1}\times P_1(i)+C_{P_2}\times P_2(i)+...+C_{P_n}\times P_n(i)}{C_{P_1}+C_{P_2}+...+C_{P_n}} & C_{P_1}+C_{P_2}+...+C_{P_n}\neq 0\\ 0 & otherwise\end{cases} \quad (2)$$

$$C_P=C_{P_1}\times DS(P_1,P)+C_{P_2}\times DS(P_2,P)+...+C_{P_n}\times DS(P_n,P) \quad (3)$$

A drawback of the model defined in [15]was that the shape of the resultant protein was independent of its constituent proteins' concentrations and this made some discontinuity problems. That is when a protein's concentration degraded, its effect in the shape of the resultant protein didn't decrease and when the concentration reached zero all of a sudden its effect vanished. However with our definition, this problem is solved and in our model proteins with higher concentration levels have larger effects on the formation of the resultant protein.

5) Protein-gene interactions

Regulatory, environment and cell receptor genes in the artificial cell has an structure shown in Fig. 1 where (RP,ATR,CTR) , (PP,ATP,CTP) , (CP,CC) define the repressor region, promoter region and coding region of the gene respectively. The behavioral genes have also the same structure, but they don't necessarily have coding region. The definition of each part is followed.

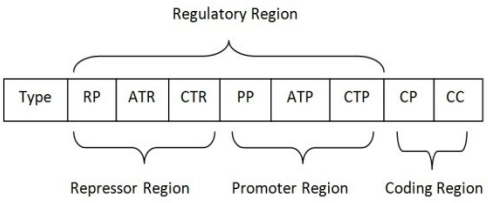

Fig. 1. Gene regions

RP : Repressor region Protein.
ATR : Affinity Threshold of Repressor region
CTR : Concentration Threshold of Repressor region
PP : Promoter region Protein.
ATP : Affinity Threshold of Promoter region
CTP : Concentration Threshold of Promoter region

The regulatory region of the gene determines its degree of activation. When the degree of similarity between the cytoplasm protein and the promoter (repressor) protein shape defined by PP (RP)is greater than ATP (ATR) and concentration level of the cytoplasm protein is also greater than CTP (CTR) then the gene is promoted (repressed) by a number in interval $[0,1]$ which we name degree of promotion (degree of repression) and denote it by DP (DR).Based on these explanations we define DP and DR as functions of cytoplasm protein shape P and concentration level C by relations (4) and (5).

$$DP(P,C)=\begin{cases}0 & DS(P,PP)<ATP\ or\ C<CTP\\ DS(P,PP)\times C & otherwise\end{cases} \quad (2)$$

$$DR(P,C)=\begin{cases}0 & DS(P,RP)<ATR\ or\ C<CTR\\ DS(P,RP)\times C & otherwise\end{cases} \quad (3)$$

The degree of activation of a gene DA is a determined by relation (6):

$$DA_{new}=\begin{cases}0 & DA_{old}+DP-DR<0\\ 1 & DA_{old}+DP-DR>1\\ DA_{old}+DP-DR & otherwise\end{cases} \quad (4)$$

In the model of [15] a gene could only be repressed or promoted. However, by adding repressor and promoter regions to all genes in our model, a gene could both be repressed and promoted depending on the shape and concentration level of the cytoplasm protein.

The gene behaviors are determined due to the following rules:

1) When a regulatory, environmental or cell receptor gene is activated, their coding region is expressed. The concentration level of the expressed protein C_{ex} would be a function of the degree of activation of the gene , DA , and the coding concentration of the gene, CC :

$$C_{ex}=DA\times CC \quad (5)$$

2) When a regulatory (environmental) gene is activated its expressed protein is released into the cytoplasm (environment of the cell) and is composed with them to form a new protein.
3) When a cell receptor gene is activated, the expressed protein interacts with the environment protein to form two proteins. Resultant protein and remainder protein. Resultant protein enters the cytoplasm and the remainder protein replaces the environment protein. The shape and the concentration level of the expressed, environment, resultant and remainder proteins are denoted by P_{env} , $P_{\exp}$, P_{res} , P_{rem} , C_{env} , $C_{\exp}$, C_{res} and C_{rem} respectively. The resultant and remainder proteins are calculated by relations (8) and (9). .

$$P_{res}(i)=P_{\exp}(i)\times P_{env}(i) \quad (6)$$

$$C_{res}=\min(C_{\exp},C_{env})$$

$$P_{rem}(i)=\begin{cases}C_{env}\times P_{env}(i)-C_{res}\times P_{res}(i) & C_{env}\times P_{env}(i)-C_{res}\times P_{res}(i)\geq 0\\ 0 & otherwise\end{cases} \quad (7)$$

$$C_{rem}=C_{env}-C_{res}\times DS(P_{env},P_{res})$$

4) Defining remainder protein is another difference of our model with the one described in [15]. In that model entering a protein to the cytoplasm from the environment didn't affect the environment proteins. However what we expect and happens in nature is that environment proteins should be lessened when a portion of them enters the cell.
5) When a behavioral gene is activated, the output defined for that gene is generated.

In biological cells the cytoplasm proteins are consumed and so their concentration degrades with time. In the artificial cell defined in this paper, we define a decay rate (dcr) so that in each iteration the concentration of the cytoplasm protein is decreased by a small percentage. Without this, the saturation of the cytoplasm protein would happen so quickly and in return would prohibit the cell from having a dynamic behavior over time.

$$C_{new}=C_{old}\times(1-dcr) \quad (8)$$

III. Case Study

In this part we use our model of artificial cell to control a two dimensional robot. In [15] a similar experiment is done.

Here comes the detailed experiment. We have a field with four obstacles placed on it and we have a robot that can move along the field in one of the four directions. Right or left and up or down. The robot's initial location is at the middle down of the field. The robot is controlled by means of an artificial cell, that is, the output of the cell is a signal indicating in which direction the robot should move and how much. At time $t=0$ the cell starts developing and without any further interrupt, it controls the robot. Our goal is to find an artificial cell with proper parameters to control the robot in a way that the robot moves along and exits the field from the top without hitting the obstacles or hitting the walls. The field with obstacles is shown in Fig. 2. The two axes represent a coordinate system for robot's position. According to this coordinate system, the robot starts its travel from point $(50,0)$ which is marked by an asterisk on Fig. 2.

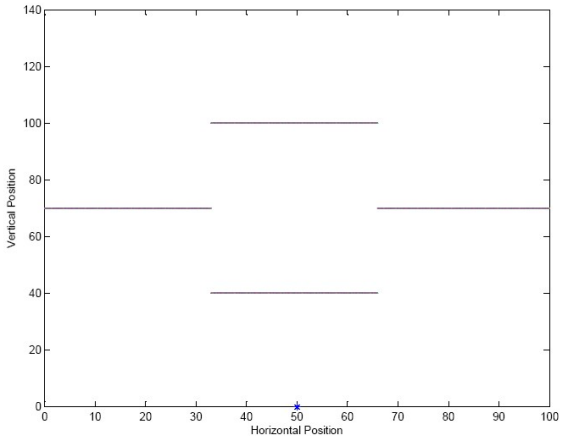

Fig. 2. The field with obstacles

Three kinds of cells are used in Genetic Algorithms(GA). The first kind has four behavioral genes, one for each of the directions and no other genes. The second kind has four behavioral genes, one for each of the directions and two extra regulatory genes. The third kind has four behavioral genes, one for each of the directions and four extra regulatory genes. The four behavioral genes in all kinds act as regulatory genes too, that is they have coding region and thus their rate of activation affects other genes' rates of activation. The proteins in all the cells are defined to be of order one which means all proteins are represented by a single value between zero and one. Collision with each obstacle has a penalty of value 10. Collision with walls has a penalty of value 50 and not being able to leave the field in a certain number of iterations has a penalty of value 200. For each kind of cell,

GA is run for 50 times. Crossover rate is set to 0.8 and mutation to 0.2 and the population size is 100. GA runs up to 150 generations. Out of these 50 runs for each kind of the cells, 13 runs of the first kind, 20 runs of the second kind, and 17 runs of the third kind could evolve a cell to move the robot to the end of the field without hitting the obstacles or the walls. The average number of generations these successful runs needed to reach the proper cell was 29 for the first kind, 32 for the second kind and 17 for the third kind. Fig. 3 shows the decrease of the objective function in one of the successful runs till it reaches zero in generation 99. Fig. 4 shows path of the robot created by one of the cells evolved through this run and Fig. 5 shows the activation degree of the four genes of this cell during its life time.

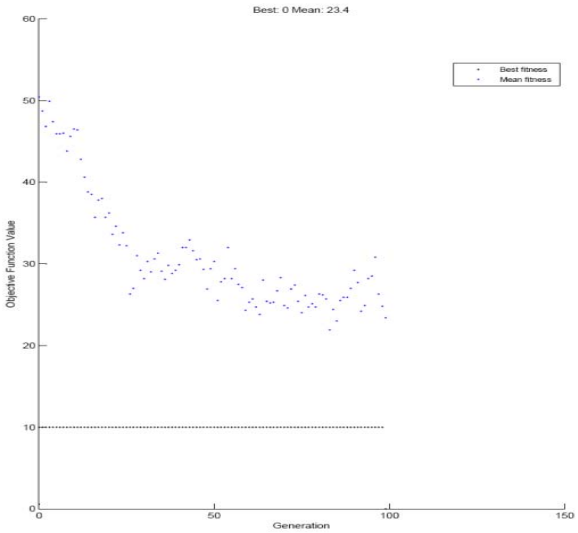

Fig. 3. Decrease in objective function during GA run

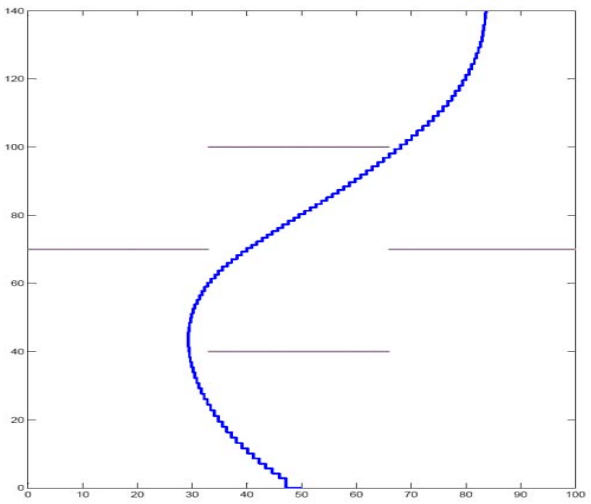

Fig. 4. Path of the robot controlled with a cell having four genes

In [15] the experiment was initialized by a cell having one environment gene, two behavioral genes and six regulatory genes using a population size of 100 and running the GA up to 500 generations and the proteins were taken to be of dimension 15×15 that is each protein was represented by a 15×15 matrix whose elements are zero or one. Fig. 6 and Fig. 7 show paths of the robot controlled by two of the proper cells found in [15].

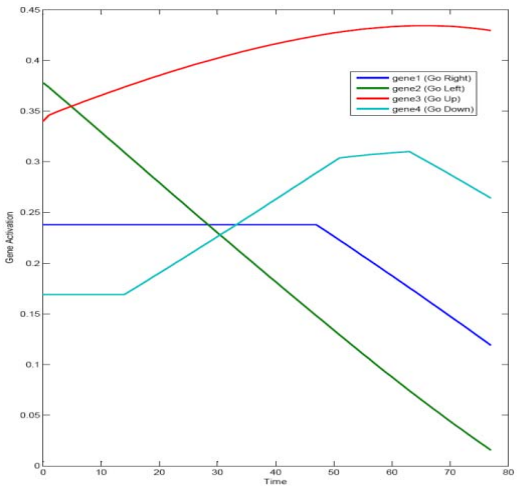

Fig. 5. Gene Activation Degree of a cell during its life time

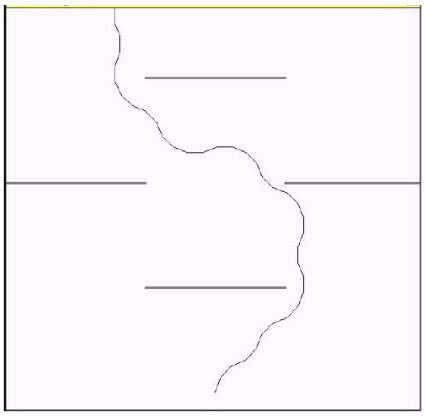

Fig. 6. Path of the robot controlled by fractal cells in [15]

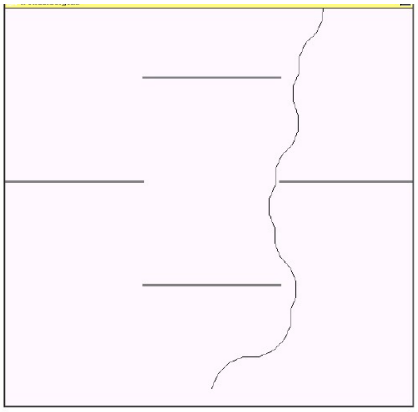

Fig. 7. Path of the robot controlled by fractal cells in [15]

In [15] the path is travelled by the robot in 32 developmental steps. However for the cells found in our experiment the number of steps is more than 100. It would be better if we could reduce the number of steps. Once the proper cell is found, the reduction of steps by n times, can be done by changing the phenotypic behavior of the cell in a way that the signals for moving the robot right or left and up or down are affected by the cell in every n iterations of running the cell. That is the corresponding signals produced by the cell in n iterations are added and affected once. Using this technique we could move the robot in six steps satisfying the problem conditions. The reduced path created by the same cell is shown in Fig. 8.

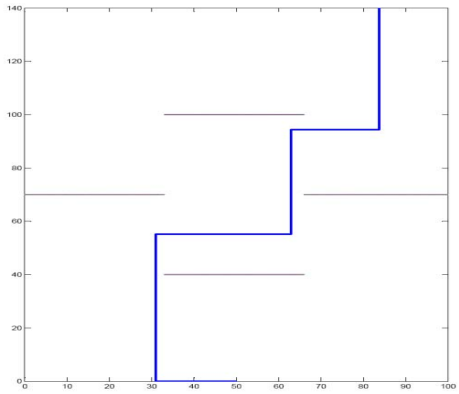

Fig. 8. Reduced step path of the robot controlled by a cell having 4 genes

A comparison of the results of our experiment with the one in [15] is presented in Table I.

TABLE I: .COMPARISON BETWEEN TWO CLASS OF ARTIFICIAL CELLS

Property	*Artificial Cells introduced in this paper*	*Artificial Fractal Cells in* [15]
Minimum number of Genes	4 genes	9 genes
Minimum Required Parameters for Genetic Algorithm	38 parameters	At least 57 parameters
Required Generations	29 generations with 100 population size	500 generations with 100 population size
Protein Dimension	1x1	15x15
Final Control Steps	6 steps	32 steps

IV. CONCLUSION

In this paper we refined and generalized previously stated artificial genetic regulatory network models. Our model can be thought as a generalization of the model in [15] in following areas:

1) The model in [15] uses subsets of Mandelbrot set in the forms of matrices whose elements are zero or one as proteins while our model uses any possible array of values between zero and one.
2) The concentration of combining proteins in the model of [15] has no effect in the shape of the resultant protein, however in our model, proteins with larger concentrations have larger effects.
3) The rules for updating the cell in each iteration (chemistry rules) have been simplified as far as possible in our model which gives us on one hand a better understanding of the cell behavior and on the other hand makes calculations less complex.
4) Our model provides a feedback passage for the artificial cell. However we plan to use this feature of our model in future works which we expect to get better results.

The first simulation results of our model compared with the ones of [15] shows that in this particular experiment our model is able to generate a proper behavior with fewer numbers of genes, parameters, protein dimensions and control steps. Our future work would be toward doing more experiments with the three models in different control problems to make a thorough comparison between them and probably determine the control cases for which each model does its best.

REFERENCES

[1] E. H. Davidson, *Genomic Regulatory Systems, in Development and Evolution*, Academic Press, 2001.
[2] W. Arthur, "The Origin of Animal Body Plans," *A Study in Evolutionary Developmental Biology*, Cambridge, 2000.
[3] I. Shmulevich, E. R. Dougherty, S. Kim, and W. Zhang, "Probabilistic Boolean Networks: a Rule Based Uncertainty Model for Gene Regulatory Networks," *Bioinformatics*, vol. 18, no. 2, pp. 261–274, 2002.
[4] I. Shmulevich, E. R. Dougherty, and W. Zhang, "From Boolean to Probabilistic Boolean Networks as Models of Genetic Regulatory Networks," in *Proc. of IEEE*, vol. 90, no. 11, pp. 1778–1792, 2002.
[5] K. Murphy and S. Mian, "Modelling Gene Expression Data using Dynamic Bayesian Networks," Tech. Rep., University of California, Berkeley, Calif, USA, 1999.
[6] R. Albert, "Boolean Modeling of Genetic Regulatory Networks," *Lecture Notes in Physics*, Springer, no. 650, pp. 459–481, 2004.
[7] T. Chen, H. L.He, and G. M. Church, "Modeling Gene Expression with Differential Equations," in *Pacific Symposium on. Biocomputing*, vol. 4, pp. 29–40, Hawaii, USA, January 1999.
[8] C. Chaouiya, E. Remy, and Thieffry, "Petri Net Modelling of Biological Regulatory Networks," *Journal of Discrete Algorithms*, vol. 6, no. 2, pp. 165-177, 2008.
[9] S. Kim, H. Li, E. R. Dougherty *et al*., "Can Markov Chain Models Mimic Biological Regulation," *Journal of Biological Systems*, vol. 10, no. 4, pp. 337–357, 2002.
[10] J. C. Bongard, "Evolving modular genetic regulatory networks," in *Proc. IEEE 2002 Congress on Evolutionary Computation*., pp. 1872-1877
[11] S. Kumar and P. J. Bentley, "Biologically inspired evolutionary development," in *Proc. 5th Int. Conf. on Evolvable Systems: From Biology to Hardware. LNAI*, Springer.
[12] P. J. Bentley, "Fractal Proteins," *Genetic programming and Evolvable Machines*, 2004.
[13] T. Quick, "Evolving Embodied Genetic Regulatory Network-Driven Control Systems," in *Proc. of ECAL*, pp. 266-277, 2003
[14] T. Taylor, "A Genetic Regulatory Network-Inspired Real Time Controller for a Group of Under Water Robots," *Intelligent Autonomous Systems*, vol. 8, 2004.
[15] P. Bentley, "Evolving Fractal Gene Regulatory Networks for Robot Control," *Lecture Notes in Computer Science, Advances in Artificial Life, Springer Berlin*, vol. 28, pp. 753-762, 2013.

Navid Rahbari Asr received his Bachelor of Science in Electrical Engineering from Tabriz University, Tabriz, Iran and his Master of Science in Electrical Engineering from Tarbiat Modares University, Tehran, Iran in 2008 and 2011 respectively. He is currently a PhD student in North Carolina State University. His research interests include control, biologically inspired computational intelligence and systems biology

Vahid Johari Majd received his B.Sc. degree in 1989 from the electrical engineering department of the University of Tehran, Iran. He then received his M.Sc. and Ph.D. degrees in the area of Control Theory from the electrical engineering department of the University of Pittsburgh, PA, U.S.A. in 1991 and 1995, respectively. He is currently an associate professor in the control system department of Tarbiat Modares University, Tehran, Iran, and is the director of intelligent control systems laboratory. His areas of interest include: Intelligent identification and control, multi-agent learning, fuzzy control, cooperative control, formation control, robust nonlinear control, fractional order control, and systems biology.