

Evolving Complex Neural Networks

Mauro Annunziato¹, Iliaria Bertini¹, Matteo De Felice², Stefano Pizzuti¹

¹ Energy, New technology and Environment Agency (ENEA)
Via Anguillarese 301, 00123 Rome, Italy

{mauro.annunziato, ilaria.bertini, stefano.pizzuti}@casaccia.enea.it

² Dipartimento di Informatica ed Automazione, Università degli Studi di Roma “Roma Tre”,
Via della Vasca Navale 79, 00146 Rome, Italy
matteo.defelice@casaccia.enea.it

Abstract. Complex networks like the scale-free model proposed by Barabasi-Albert are observed in many biological systems and the application of this topology to artificial neural network leads to interesting considerations. In this paper, we present a preliminary study on how to evolve neural networks with complex topologies. This approach is utilized in the problem of modeling a chemical process with the presence of unknown inputs (disturbance). The evolutionary algorithm we use considers an initial population of individuals with different scale-free networks in the genotype and at the end of the algorithm we observe and analyze the topology of networks with the best performances. Experimentation on modeling a complex chemical process shows that performances of networks with complex topology are similar to the feed-forward ones but the analysis of the topology of the most performing networks leads to the conclusion that the distribution of input node information affects the network performance (modeling capability).

Keywords: artificial life, complex networks, neural networks

1 Introduction

Artificial Neural Networks (ANN) and Evolutionary Algorithms (EA) are both abstractions of natural processes. They are formulated into a computational model so that the learning power of neural networks and adaptive capabilities of evolutionary processes can be harnessed in an artificial life environment. “Adaptive learning”, as it is called, produces results that demonstrate how complex and purposeful behavior can be induced in a system by randomly varying the topology and the rules governing the system. Evolutionary algorithms can help determine optimized neural network architectures giving rise to a new branch of ANN known as Evolutionary Neural Networks [1] (ENN). It has been found [2] that, in most cases, the combinations of evolutionary algorithms and neural nets perform equally well (in terms of accuracy) and were as accurate as hand-designed neural networks trained with backpropagation [3]. However, some combinations of EAs and ANNs performed much better for some data than the hand-designed networks or other EA/ANN combinations. This suggests that in applications where accuracy is a premium, it might pay off to experiment with

EA and ANN combinations. A new and very interesting research area which recently emerged is that of *Complex Networks* (CN). CN (mainly *scale-free* networks) are receiving great attention in the physics community, because they seem to be the basic structure of many natural and artificial networks like proteins, metabolism, species network, biological networks [4, 5, 6], the Internet, the WWW, the e-mail network, metabolic networks, trust network and many more [7].

In this context, using complex ANN topologies driven by evolutionary mechanisms is a new idea and we used them in order to model complex processes.

2 The methodology

In this context, the goal of the proposed work is the study of evolutionary neural networks with a directed-graph based topology, obtained using an iterative algorithm similar to that proposed by Barabasi-Albert in 1999 [8].

2.1 complex networks

A unique definition of “complex network” doesn’t exist, this term refers to networks with non-trivial topology and high number of nodes and connections. However, complex networks can be classified, according to some topology descriptors, into two main classes : *Small World* and *Scale Free*.

The most important topology descriptors are: the node degree distribution, the shortest path length and the clustering coefficient.

Properties of these networks are often compared with random graphs [9] that are to be considered “simple” networks. Random networks have a Poisson node degree distribution, a small shortest path length and a small clustering coefficient.

Small World networks [4, 10] have a Poisson node degree distribution, a small shortest path length and a high clustering coefficient. They are in the middle between regular and random networks (see Figure 1) and it has been shown [4] that this topology is the optimal one for communication tasks.

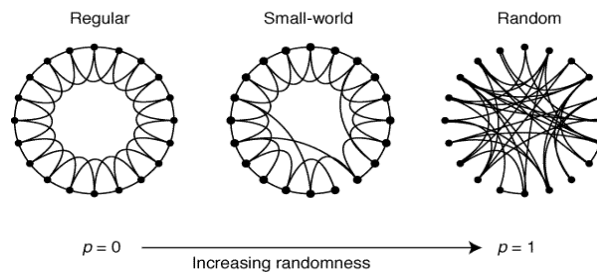


Figure 1 : networks topologies

The scale-free model [11] have a node degree distribution which follows the power law distribution. It means that we have few nodes with high connectivity (*hubs*) and many nodes with few links (see Figure 2). These networks show the so-called “small-world” property [4], every two nodes of the network are placed at a distance of a relative small number of edges. These types of networks are receiving great attention

in the physics community, because many networks have been reported recently to follow a scale free degree distribution. Just as examples we can cite the Internet, the WWW, the e-mail network, metabolic networks, trust network and many more [7]. Their inspiring philosophy could be synthesized in the sentence “the rich gets richer”, because each node has a probability to get a new link that is proportional to the number of its current links.

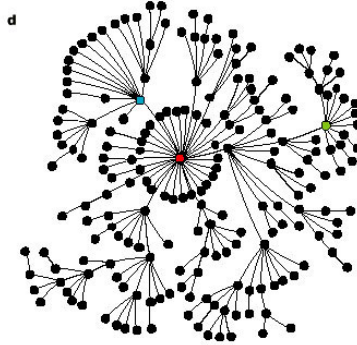


Figure 2: scale free network

In this study, we focussed on scale-free networks.

2.2 The algorithm

We used ANNs based on a complex network created with the algorithm whose pseudo-code is shown in Table 1. The algorithm starts creating an initial set of nodes connected each other and then each added node is connected with a selected destination node with a *preferential-attachment* function: this function defines the probability that a node in network receive a link from a newly inserted node [8, 12]. The analytic form of this function is:

$$\Pi(k_i) = \frac{k_i^\alpha}{\sum_j k_j^\alpha} \quad (1)$$

In this function k_i is the degree of node i . This function is monotonically increasing, the α parameter influences the numbers of dominant hubs with high connectivity. In figures 3 e 4 we show the node degree distributions of networks with 4000 nodes built with the presented algorithm with different α values, fitting with a power function like $k^{-\gamma}$ is shown in following plots (figures 3, 4).

The output node of the network is randomly chosen after the insertion of the nodes of the hidden layer. At the selection of the output node, input nodes are inserted.

Table 1: algorithm pseudo-code

```

BEGIN
  /* Initial set of nodes */
  FOR i = 1 to  $m_0$ 
    ADD nodei
    CONNECT nodei to ALL
  END FOR
  /* Add nodes and connect them with PA function */
  FOR i = 1 to TOTAL_NODES
    ADD nodei
    FOR j = 1 to m
      x = GET_NODE_WITH_PREFERENTIAL_ATTACHMENT
      CONNECT nodei to nodex
      CONNECT nodex to nodei
    END FOR
  END FOR
  /* Select output node */
  x = RANDOM(TOTAL_NODES)
  OUTPUT_NODE = nodex
  /* Add and connect input nodes */
  CASE INPUT_CONNECTION_TYPE OF:
    /* CONNECTION TYPE A */
    A: ADD ALL_INPUT_NODES
    FOR i = 1 to m
      x = RANDOM(TOTAL_NODES)
      CONNECT ALL_INPUT_NODES to nodex
    END FOR
    /* CONNECTION TYPE B */
    B: FOR i = 1 to TOTAL_INPUT_NODES
      ADD input_nodei
      FOR j = 1 to m
        x = GET_NODE_WITH_PREFERENTIAL_ATTACHMENT
        CONNECT input_nodei to nodex
      END FOR
    END FOR
  END CASE
END

```

In this algorithm we considered two types of connections between the network and the input nodes (in the pseudo-code these methods are indicated by the variable INPUT_CONNECTION_TYPE). In case A we have all the input nodes connected to the same m nodes of the hidden layer. Otherwise in case B we have each input node linked to m random nodes of the hidden layer.

A network created with this algorithm is presented in Figure 5.

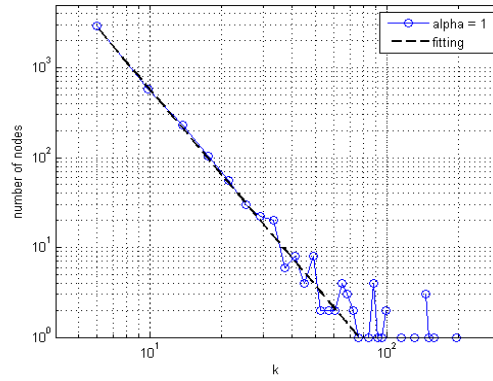


Figure 3: degree distribution of a 4000 nodes network created with $\alpha = 1$ fitted with a power-law function with $\gamma = 3.1$ (dashed line)

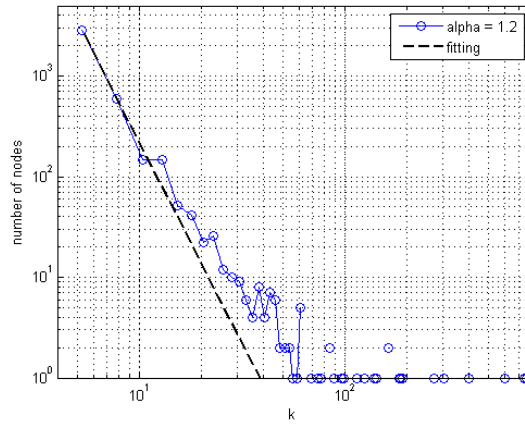


Figure 4: degree distribution of a 4000 nodes network created with $\alpha = 1.2$ fitted with a power-law function with $\gamma = 4$ (dashed line)

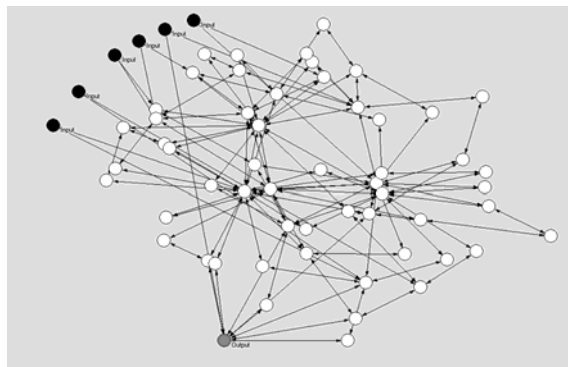


Figure 5: an example of complex neural network. Input nodes are black and the output node is grey.

We used the values of the parameters presented in the Table 2 for the execution of the algorithm.

2.3 The evolutionary environment

The implemented evolutionary environment is an *Artificial Life* (ALIFE) environment [13]. This approach has been tested on the optimization of static well known benchmarks, as the Travelling Salesman Problem, the Chua's circuit and the Kuramoto dynamical system [14], as well as real cases [15, 16, 17, 18]. The ALIFE context is a two-dimensional lattice (life space) representing a flat physical space where the artificial individuals (or autonomous agents) can move around. At each iteration (or life cycle), individuals move in the life space and, in case of meeting with other individuals, interaction occurs. Each individual has a particular set of rules that determines its interactions with other agents basically based on a competition for energy in relation to the performance value. Individuals can self-reproduce via haploid mutation which occurs only if the individual has an energy greater than a specific birth energy. In fact, during reproduction, an amount of energy equal to the birth energy is transferred from the parent to the child. In the haploid reproduction a probabilistic test for self reproduction is performed at every life cycle and a probabilistic-random mutation occurs on the genes according to the mutation rate and the mutation amplitude, which are evolutionary themselves [19]. When two individuals meet, fight occurs. The winner is the individual characterized by a greater value of performance and the loser transfers an amount of energy (fighting energy) to the winner. At every life cycle each individual age is increased and when the age reaches a value close to the average lifetime, the probability of natural death increases. This ageing mechanism is very important to warrant the possibility to lose memory of very old solutions and follow the process evolution. Another mechanism of death occurs when the individual reaches the null energy due to reproduction or fighting with other individuals. For interested readers, a detailed description of the methodology is reported in [15, 16]. The characteristic Artificial Life environment we used is called "Artificial Societies", introduced in [20], but we made some modifications to the original one. In our implementation each individual of the initial population is initialized with a different network topology in his genotype. Initial networks are created with a value of the parameter m that varies from 2 to 6. Network weights and activation functions, but not the topology, are subject to random mutations and no crossover mechanism among different network topologies has been implemented because in the artificial life algorithm we used there is not present a bisexual reproduction mechanism.

Table 2: parameters of the algorithm

m_0	4
m	2-6
α	1.2
Life space dimension	25 x 25
Initial population	215

3 The benchmark model

The model we used is a process consisting of two linearized *Continuous flow Stirred Tank Reactor* (CSTR) models in parallel (figure 4). The CSTR model describes an exothermic diabatic reaction of the first order and it is commonly studied for his characteristics [21, 22]. The equations, written in dimensionless form [23], are the following:

$$\begin{aligned}\dot{x}_1 &= q(x_{1s} - x_1) - \Phi x_1 \kappa(x_2) \\ \dot{x}_2 &= q(x_{2s} - x_2) - \delta(x_2 - x_3) + \beta \Phi \kappa(x_2) \\ \dot{x}_3 &= \delta_1 [q_c(x_{3s} - x_3) + \delta \delta_2 (x_2 - x_3)]\end{aligned}\quad (2)$$

where x_1 , x_2 and x_3 are respectively dimensionless concentration, reaction temperature and cooling-jacket temperature. The first manipulated input is q_c which represents the cooling-jacket flow rate and the second input q is the reactor flow rate. The output of the system is the dimensionless concentration x_1 . The other parameters are explained in Table 3 with the values we set. The model we used is represented in the schema in figure 6. As stated before, the process consists of two CSTR linearized models with different Damkholer numbers, therefore we have two different inertias (the length of the transient regime), one fast and one slow. In order to simulate a non-stationary environment (like most of real situations are), we consider the input q (reactor flow rate) of the CSTR model as “disturbance”, because it is not given to the neural models. The whole output of the system, $s(t)$, is the sum of the two CSTR’s output normalized between 0 and 1 and sampled every 0,3 s.

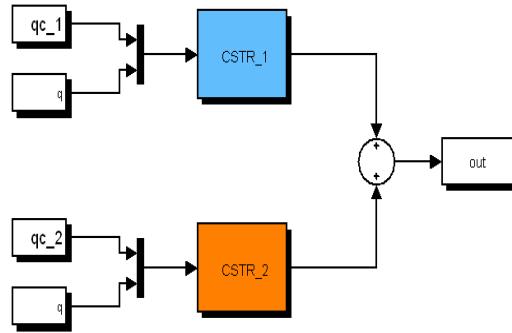


Figure 6: model layout

The system matrices of the model with “disturbance” are the following:

$$A = \begin{bmatrix} -q_s - \Phi K & -\Phi x_{1s} K' & 0 \\ \beta \Phi K & -q_s - \delta - \beta \Phi x_{1s} K' & \delta \\ 0 & \delta \delta_1 \delta_2 & -\delta_1 q_{cs} - \delta \delta_1 \delta_2 \end{bmatrix}\quad (3)$$

$$B = \begin{bmatrix} 0 & (x_{1f} - x_{1s}) \\ 0 & (x_{2f} - x_{2s}) \\ \delta_1 (x_{3f} - x_{3s}) & 0 \end{bmatrix}\quad (4)$$

$$\text{With } K = e^{\frac{x_{2s}}{1+\frac{x_{2s}}{\gamma}}} \text{ and } K' = e^{\frac{x_{2s}}{1+\frac{x_{2s}}{\gamma}} - \frac{1}{\left(1+\frac{x_{2s}}{\gamma}\right)^2}}.$$

The reactor flow rate q (disturbance) is modeled as a train pulse of 0.1 Hz of frequency, 0.15 of amplitude and a pulse width of 40%.

Both CSTR have in input a random step ranging between 0 and 0.5 every 6.5 seconds. The step is filtered by a $\frac{1}{s+1}$ transfer function.

Table 3: model parameters

Name	Value	Explanation
x_{1s}	0.2028	Steady state of concentration
x_{2s}	5	Steady state of reaction temperature
x_{3s}	0.4079	Steady state of cooling-jacket temperature
x_{1f}	1	Dimensionless reactor feed concentration
x_{2f}	0	Dimensionless reactor feed temperature
x_{3f}	-1	Dimensionless cooling-jacket feed temperature
Φ	0.25-0.53	Damkholer number
q_{cs}	0.9785	Steady state of cooling-jacket flow rate
q_s	1	Steady state of reactor flow rate
β	8	Dimensionless heat of reaction
δ	0.3	Dimensionless heat transfer coefficient
δ_1	10	Reactor to cooling jacket volume ratio
δ_2	1	Reactor to cooling jacket density heat capacity ratio
γ	20	Dimensionless activation energy

The term $k(\cdot)$ represents the dimensionless Arrhenius reaction rate which corresponds to:

$$\kappa(x) = e^{\frac{x}{1+\frac{x}{\gamma}}} \quad (5)$$

4 Experimental results

The output of the network is the prediction of the signal $s(t+h)$, with h the prediction horizon, and the inputs are the six past samples of the signal, $s(t-1)\dots s(t-6)$. All the experimentations are carried out with the prediction horizon (h) as 5.

We made a set of 10 tests and calculated the mean RMSE :

$$E_{rmse} = \sqrt{\frac{\frac{1}{2} \sum_{i=1}^M (y(i) - \bar{y}(i))^2}{M}} \quad (6)$$

In Table 4 it is presented a comparison of the performance of different ANN typologies: feed-forward network, fully-connected network and scale free networks trained with ALIFE algorithm and feed-forward network trained with classic back-propagation algorithm. An example of target signal and neural network output is show in figure 7.

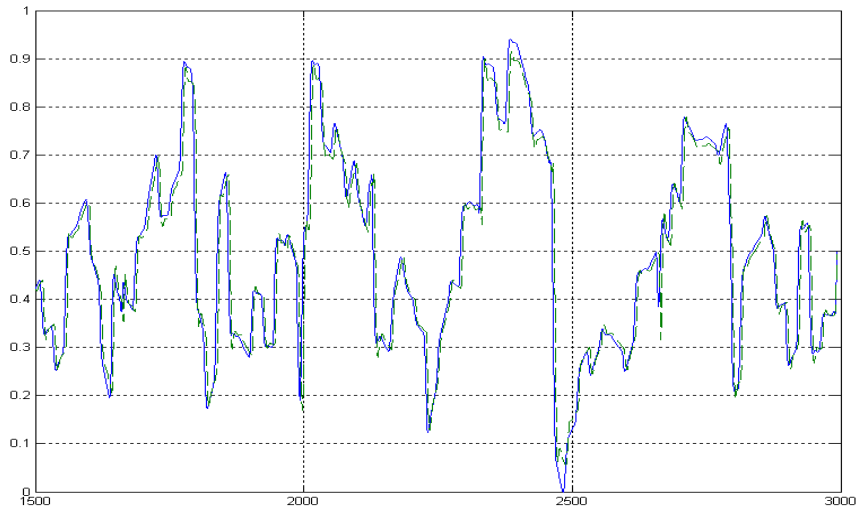


Figure 7: comparison between the model output and the output of the ANN (dotted line)

Table 4: performance comparison

ANN typology	RMSE
Feed-forward network with ALIFE	0.039
Feed-forward network with back-propagation	0.072
Fully-connected with ALIFE	0.039
ANN with scale-free topology with ALIFE	0.041

At the end of the tests a topology analysis on the “best” networks was performed. In fact, natural selection mechanism, intrinsic in evolutive algorithms, generates a drastic reduction of the number of different topologies which are inside the solutions of the overall population. An example of this mechanism is shown in figure 8: on the

horizontal axis time it is reported, while on the vertical axis there are the different network topologies. In this plot it is clearly shown the process leading to the elimination of most network topologies.

An analysis of the average connectivity, the m parameter of tables 1 and 2, of the “best” networks shows that, at the end of 10 simulations, the value was 3.9, an outcome very close to the arithmetic mean (the m parameter ranges from 2 to 6). In this way, it seems that connectivity doesn’t affect the performance of the neural network.

A network analysis considering parameters like clustering coefficient or shortest path length has not been performed because of the small dimension of the networks considered, in fact with small networks some parameters commonly used in complex networks analysis are to be considered not meaningful.

Moreover, we analyzed the type of input nodes connection and we observed that the 80% of the “best” networks had used an input nodes connection of type B (see variable `INPUT_CONNECTION_TYPE` in the pseudo-code presented in Table 1), so after these tests we can see that input distributed information leads to a better performance than an information focussed on little sets of nodes.



Figure 8: example of natural selection mechanism

5 Conclusion

Complex networks like the scale-free model proposed by Barabasi-Albert are observed in many biological systems and the application of this topologies to artificial neural network leads to interesting considerations.

In this paper, we presented a preliminary study on how to evolve neural networks with complex topologies and in particular we focused on the scale-free model. As benchmark we faced the problem of modeling a chemical process with the presence of unknown inputs (disturbance). The evolutionary environment has an initial population

of individuals with different scale-free networks in the genotype and at the end of the algorithm we observed and analyzed the topologies of networks with the best performances.

The experimentation we did is to be considered only the beginning of the exploration of the union between neural networks and complex topologies. It is necessary to perform more extended tests and to compare the structures object of this paper with another optimization and modeling methods like kernel machines (SVRs).

Testing showed that performances of complex networks on the proposed modeling problem are similar to those achieved with classic feed-forward networks. The possibility to perform topological analysis at the end of the evolutionary process, typical of the algorithms like that one we used, could be considered as the basis for a profitably union between the two paradigms of cognitive systems par excellence: artificial neural networks and complex networks. In fact, the analysis at the end of a set of tests, showed on one hand that connectivity has little influence on the network performance, while on the other hand the input connectivity seems to be more effective. These results suggest a direction to the purpose of creating networks that are able to accomplish complex modeling tasks.

References

1. Yao, X., "Evolving Artificial Neural Networks", Proceedings of the IEEE, 87(9): (1999)1423-1447
2. Alander, J. T., "An indexed bibliography of genetic algorithms and neural networks", Technical Report 94-1-NN, University of Vaasa, Department of Information Technology and Production Economics (1998)
3. Cant-Paz, E. and Kamath, C. , "An empirical comparison of combinations of evolutionary algorithms and neural networks for classification problems", IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics (2005) 915-927.
4. Watts DJ, Strogatz SH., "Collective dynamics of 'small-world' networks", Nature. 1998 Jun 4;393(6684):440-2.
5. Gully A. P. C. Burns, Malcolm P. Y Oung, "Analysis of the connectional organization of neural systems associated with the hippocampus in rats", Philosophical Transactions of the Royal Society B: Biological Sciences, Volume 355, Issue 1393, 29 Jan 2000, Pages 55 – 70
6. Victor M. Eguiluz, Dante R. Chialvo, Guillermo A. Cecchi, Marwan Baliki, and A. Vania Apkarian, "Scale-Free Brain Functional Networks", Phys. Rev. Lett. 94, 018102 (2005)
7. Dorogotvsev, S. N., Mendes, J. F. F., "Evolution of Networks", Advances in Physics, Vol. 51, n. 4, pp. 1079-1187, 2002
8. Barabasi, A.-L. and Albert R. 1999, "Emergence of scaling in random networks", Science, 286 509–512.
9. P Erdos, A Renyi, "On random graphs", Publ. Math. Debrecen, 1959
10. Watts, D. J., "Small Worlds: The Dynamics of Networks Between Order and Randomness", Princeton University Press, 1999

11. Coven, R., Havlin, S., ben-Avraham, D., "Structural Properties of Scale-Free Networks", chap. 4 in "Handbook of graphs and networks", Eds. S. Bornholdt and H. G. Schuster, Wiley-VCH, 2002
12. H. Jeong, Z. Neda and A.-L. Barabasi, "Measuring preferential attachment for evolving networks", Euro. Phys. Lett. 61 567 (2003)
13. Langton, C., "Artificial Life", Addison-Wesley, Redwood City/CA, USA, 1989
14. Annunziato M., Bertini I., Lucchetti M., Pannicelli A., Pizzuti S. "Adaptivity of Artificial Life Environment for On-Line Optimization of Evolving Dynamical Systems", in Proc. EUNITE01, Tenerife, Spain, 2001
15. Annunziato M. , Bertini I. , Pannicelli A., Pizzuti S. , Tsimring L., "Complexity and Control of Combustion Processes in Industry", Proc. of CCSI 2000 Complexity and Complex System in Industry, Warwick, UK, 2000
16. Annunziato M. , Lucchetti M., Orsini G., Pizzuti S. , "Artificial life and on-line flows optimisation in energy networks", IEEE Swarm Intelligence Symposium, Pasadena (CA), USA, 2005
17. Annunziato M., Bertini I., Pannicelli A. and Pizzuti S. "A Nature-inspired-Modeling-Optimization-Control system applied to a waste incinerator plant", 2nd European Symposium NiSIS'06, Puerto de la Cruz, Tenerife (Spain), 2006
18. Annunziato M., Bertini I., Pannicelli A., Pizzuti S. "Evolutionary Control and On-Line Optimization of an MSWC Energy Process", Journal of Systemics, Cybernetics and Informatics, Vol.4, Num. 4, 2006
19. Annunziato M., Bertini I., Iannone R., Pizzuti S. "Evolving feed-forward neural networks through evolutionary mutation parameters", 7th International Conference of Intelligent Data Engineering and Automated Learning (IDEAL 06), Burgos, Spain, 2006, 554-561
20. Annunziato M. , Bruni C., Lucchetti M. , Pizzuti S. : "Artificial life approach for continuous optimisation of non stationary dynamical systems", Integrated Computer-Aided Engineering, vol. 10, n.2, 111-125, June 2003
21. Russo L.P., Bequette B.W. "Impact of process design on the multiplicity behaviour of a jacketed exothermic CSTR", AIChE Journal, Vol. 41, 1995, pp 135-147
22. Saraf V.S., Bequette B.W., "Auto-tuning of cascade controlled open-loop unstable reactors", American Control Conference, 2001. Proceedings of the 2001, Vol.3, Iss., 2001, pp. 2021-2026
23. Russo, L.P. and Bequette, B.W., "State-Space versus Input/Output Representations for Cascade Control of Unstable Systems", Ind. Eng. Chem. Res., 36, 6, 1997, 2271 – 2278