Gene Regulation Network Use for Information Processing

Enrique Fernandez-Blanco
University of A Coruña, Spain

J. Andrés Serantes
University of A Coruña, Spain

INTRODUCTION

From the unicellular to the more complex pluricellular organism needs to process the signals from its environment to survive. The computation science has already observed, that fact could be demonstrated remembering the artificial neural networks (ANN). This computation tool is based on the nervous system of the animals, but not only the nervous cells process information in an organism. Every cell has to process the development and functioning plan encoded at its DNA and every one of these cells executes this program in parallel with the others. Another interesting characteristic of natural cells is that they form systems that are tolerant to partial failures: small errors do not induce a global collapse of the system.

The present work proposes a model that is based on DNA information processing, but adapting it to general information processing. This model can be based on a set of techniques called Artificial Embryogeny (Stanley K. & Miikkulainen R. 2003) which adapts characteristics from the biological cells to solve different problems.

BACKGROUND

The Evolutionary Computation (EC) field has given rise to a set of models that are grouped under the name of Artificial Embryology (AE), first introduced by Stanley and Miikkulainnen (Stanley K. & Miikkulainen R. 2003). This group refers to all the models that try to apply certain characteristics of biological embryonic cells to computer problem solving, i.c. self-organisation, failure tolerance, and parallel information processing.

The work on AE has two points of view. On the one hand can be found the grammatical models based on L-systems (Lindenmayer A. 1968) which do a top-down approach to the problem. On the other hand can be found the chemical models based on the Turing's ideas (Turing A. 1952) which do a down-top approach.

The grammatically approach, some times, has used the models for study the evolution of ANN, which is known as neuroevolution. The first neuroevolution system was development by Kitano (Kitano, H. 1990). In this work Kitano shows that it was possible to evolve the connectivity matrix of ANN through a set of rewrite rules. Another remarkable work is the application of L-systems do by Hornby and Pollack (Hornby, G. S. & Pollack J. B. 2002). At this work they simultaneously evolved the body morphologies and the neural network of artificial creatures in a simulated 3D physical environment. Finally, mention the works carry out by Gruau (Gruau F. 1994) where the author uses grammar trees to encode steps in the development of a neural network from a single antecesor cell.

On the chemical approach, the starting point of this field can be found in the modelling of gene regulatory networks, performed by Kauffmann in 1969 (Kauffman S.A. 1969). After that, several works were carried out on subjects such as the complex behaviour generated by the fact that the differential expression of certain genes has a cascade influence on the expressions of others (Mjolsness E., Sharp D.H., & Reinitz J. 1995). Considering the gene regulatory networks works, the most relevant models are the following: the Kumar and Bentley model (Kumar S. & Bentley P.J 2003), which uses the theory of fractal proteins Bentley, P.J., Kumar, S. 1999; for the calculation of protein concentration; the Eggenberger model (Eggenberger P. 1996), which uses the concepts of cellular differentiation and cellular movement to determine cell connections; and the work of Dellaert and Beer (Dellaert F. & Beer R.D. 1996), who propose a model that incorporates the idea of biological operons to control the model expression, where the function assumes the mathematical meaning of a Boolean function.

GENETIC REGULATORY NETWORK MODEL

The cells of a biological system are mainly determined by the DNA strand, the genes, and the proteins contained by the cytoplasm. The DNA is the structure that holds the gene-encoded information that is needed for the development of the system. The genes are activated or transcribed thanks to the protein shaped-information that exists in the cytoplasm, and consist of two main parts: the sequence, which identifies the protein that will be generated if the gene is transcribed, and the promoter, which identifies the proteins that are needed for gene transcription.

Another remarkable aspect of biological genes is the difference between constitutive genes and regulating genes. The latter are transcribed only when the proteins identified in the promoter part are present. The constitutive genes are always transcribed, unless inhibited by the presence of the proteins identified in the promoter part, acting then as gene oppressors.

The present work has tried to partially model this structure with the aim of fitting some of its abilities into a computational model; in this way, the system would have a structure similar that is similar to the above and will be detailed in the next section.

Proposed Model

Various model variants were developed on the basis of biological concepts. The proposed artificial cellular system is based on the interaction of artificial cells by means of messages that are called proteins. These cells can divide themselves, die, or generate proteins that will act as messages for themselves as well as for neighbour cells.

The system is supposed to express a global behaviour towards the information processing. Such behaviour would emerge from the information encoded in a set of variables of the cell that, in analogy with the biological cells, will be named genes.

The central element of our model is the artificial cell. Every cell has a binary string-encoded information for the regulation of its functioning. Following the biological analogy, this string will be called DNA. The cell also has a structure for the storage and management of the proteins generated by the own cell and those received from neighbourhood cells; following the biological model, this structure is called cytoplasm.

The DNA of the artificial cell consists of functional units that are called genes. Each gene encodes a protein or message (produced by the gene). The structure of a gene has four parts (see **Figure 1**):

- Sequence: the binary string that corresponds to the protein that encodes the gene
- Promoters: is the gene area that indicates the proteins that are needed for the gene's transcription.
- Constituent: this bit identifies if the gene is constituent or regulating
- Activation percentage (binary value): the percentage of minimal concentration of promoters proteins inside the cell that causes the transcription of the gene.

The transcription of the encoded protein occurs when the promoters of the non-constituent genes appear in a certain rate at the cellular cytoplasm. On the other hand, the constituent genes are expressed until such expression is inhibited by the present rate of the promoter genes.

The other fundamental element for keeping and managing the proteins that are received or produced by the artificial cell is the cytoplasm. The stored proteins have a certain life time before they are erased. The cytoplasm checks which and how many proteins are needed for the cell to activate the DNA genes, and as such responds to all the cellular requirements for the concentration of a given type of protein. The cytoplasm also extracts the proteins from the structure in case they are needed for a gene transcription.

Figure 1. Structure of a system gene