Artificial Cells for Information Processing: Iris Classification

Enrique Fernandez-Blanco, Julian Dorado, Jose A. Serantes, Daniel Rivero, and Juan R. Rabuñal
University of A Coruña, Campus Elviña, A Coruña, Spain
{efernandez, julian, jserantesv, drivero, juanra}@udc.es

Abstract. This paper presents a model in the Artificial Embryogene (AE) framework. The presented system tries to model the main functions of the biological cell model. The main part of this paper describes the Gene Regulatory Network (GRN) model, which has a similar processing information capacity as Boole’s Algebra. This paper also describes how to use it to perform the Iris Classification problem which is a pattern classification problem. The aim of this work is to show that the model can solve this kind of problems.

Keywords: Artificial Embryogeny, Genetic Algorithms.

1 Introduction

Nature presents a lot of different systems that can be used in Computer Science as an inspiration to develop new tools. Examples like Artificial Neural Networks (ANNs) or Genetic Algorithms are well-known. This work starts from the idea that any cell of a biological tissue has to communicate and process the signals of its environment. This behavior can be seen as distributed computation, where each cell plays the character of a single processor and it has to coordinate its computation with its neighbor cells. Nature just needs a few signals from the environment and the information contained in the DNA to develop and coordinate the most complicated structures.

The objective of the present work is to develop a model inspired in embryological cells, which have features like self-organization, self-reparation, etc. To develop the computer adaptation, the biological model was simplified by identifying and modeling the essential elements. In this way, some parts of the computer adaptation have very similar functions to the biological ones (DNA, gene or cytoplasm, etc.). The main objective of this paper is to adapt this model in order to apply it to information processing problems and, in particular, to classification and pattern recognition problems.

2 Background

In 2003, Stanley and Miikmulainen [1] developed a methodology to classify the different AE models that appear in Evolutionary Computation (EC). These models are
based on abstractions of the embryological cells, which can be classified into two main types. On one hand, some works follow a grammatical approach, where the most important works are related to L-systems [2]. On the other hand, other studies have a chemical-oriented approach based on Turing’s ideas [3].

The works related to the grammatical approach have been mostly used to develop ANN. Kitano’s work [4] shows how the connectivity matrix of an ANN is evolved with a set of rules. Another remarkable work is [5], in which the authors develop both the control and the structure of the robot using L-systems.

For the chemical approach, the first work to be mentioned is [6], in which Kauffmann develops his Gene Regulatory Networks [6]. The objective of the works which follow this approach is usually its application to different problems, such as approximating a simple figure/structure in a 3D space, or the development of evolutionary hardware [7]. One of the most important works which tries to solve the previously mentioned problems is described in [8]. The most interesting part is the usage of fractal proteins for the communication among the cells of the model.

The model presented in the current paper can be included into the chemical approach. The model has included most of the concepts present in the previously mentioned works, like the concept of operon or the cellular division and death. The most novel concept is that never before a chemical approach has been used to solve an information processing problem.

3 Model

Every cell of any biological tissue has as antecessor: a unique cell, called zygote, which generates other cells and they can coordinate their behavior using the information present in DNA. Each cell knows its purpose from its DNA and the proteins that it receives. Therefore, it can be considered that each cell of the tissue works as a processor and all of them operate with proteins using the DNA as operator set. Self-reparation, self-organization and parallel information processing are some features of the structures generated with this computation model [9].

Below, the structures of the artificial model that arise from the study of the biological issues are explained.

Protein

Protein is the basic piece of information. In this model, proteins are a string of bits that identifies each one of the different proteins and has a time to live (TTL). Due to this, the system has a memory of previous generated proteins, until they are used or degraded.

Cytoplasm

Cytoplasm is the part of the artificial cell which has the responsibility of managing the information inside the cell. The responsibility of this part is to manage the proteins needed for a transcription in the cell and to check the concentration level of the proteins inside and outside the cell to decide which proteins will be communicated.

Gene

Each gene of the system represents a rule, where some conditions have to be fulfilled to perform a certain computation or process. The genes are strings of bits which contain two parts: promoters and a gene identifier.