

A Discussion Outline on Gene-Regulated Development, Modularity and EC

Piet van Remortel
Computational Modeling Lab (CoMo)
Department of Computer Science (DINF)
Vrije Universiteit Brussel (VUB)
Brussels, Belgium
pvremort@vub.ac.be

Abstract

Bio-inspired development is often proposed as a mechanism that can exploit hierarchy and module reuse to alleviate the scalability problem in Evolutionary Computation. We discuss some concepts of modularisation and module decomposition in a simple setup consisting of a GA combined with a gene-regulated development mapping. By means of the concept of activation propagation, we discuss modules that are inherent to the mechanism of gene-regulated development and discuss possible ways to map such modules on the concepts or module decomposability in the total loop from genome, over development, to phenotype and associated fitness value.

1 Introduction

The lack of inherent scalability of Evolutionary Computation (EC) is a well known problem. Mechanisms that allow for module re-use in several levels of hierarchy seem to provide a possible solution to this problem. An example of such a mechanism is a bio-inspired genotype to phenotype mapping. The primary focus of our work is to investigate gene-regulated development from genotype to phenotype as a possible solution of this scaling problem. More specifically, we aim to clarify how EC and development can be combined to efficiently work together. This rather suspicious attitude is based on the apparent mismatch between the need of EC for low epistasis, and the interesting scaling properties of development which are based on high gene interaction.

Development is an interesting subject for further investigation, especially within the scope of this work-

shop, since it allows for reuse and hierarchy based on mechanisms that can be observed in action in real organisms in nature. As such, new ideas in this area can be related to biological knowledge of development, and criticised as such.

2 Development in the Evolutionary Loop

For our investigation, we use a simple evolutionary setup, where the genotype-phenotype mapping is performed by a gene-regulated development. This mechanism allows evolved genotypes to develop into the specification of a two dimensional non-uniform cellular automaton (2D CA). As a result, genotypes that are selected are used as generative plan for the development algorithm, and result in a 2D pattern of rules for the CA cells. From this CA the fitness value is derived. This can be done either by interpreting the CA as a static pattern, in which each rule-type corresponds to a different unit in the pattern, or by updating the CA over a number of time-steps and deriving the fitness value from its behaviour or final state. An overview of the setup is shown in Figure 1.

3 A Model of Development

The development model that is used in this paper is an adapted version of a model presented earlier [3], which was inspired by 'Cell Systems' [2]. The goal of our model is to implement the basic mechanisms of development while allowing detailed analysis of effects and events during development. The model has the following properties :

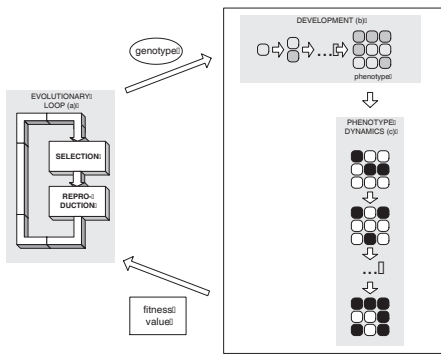


Figure 1. Overview of our experimental setup, with the complete loop from reproduction and selection, through development, into phenotype. Fitness can be derived from the phenotype by interpreting it statically of dynamically.

- Phenotypes consists of several cells, each with their own copy of a shared genome
- Each cell has its own concentrations of a variable number of proteins, along with a cell age, and a concept of differentiation. Differentiation in this case is related to the definition of transition rule of a CA cell that resides in every cell.
- The cell variables (age, proteins, ...) can result in gene activation or inhibition within each cell, implementing a simple mechanism of gene regulation. Note that the *amplitude* of the gene activation is not regulated. This means that a gene is either active, or inactive, there is no in-between.
- Two main mechanisms are implemented by which cells can change properties :
 1. Cell divisions, which can be symmetric or asymmetric. Asymmetric cell divisions are a source of creating new cell types based on the **cell lineage**. Combined with a simple implementation of nuclear determinants, this mechanism is the basis of *mosaic development* [5].
 2. Cells can communicate locally with their neighbours, which implements **cell induction**. This mechanism is the basis of *regulative development* [5].

A total developmental run consists of a predefined number of developmental steps, which is fixed for the entire experiment. During one such step, the shared

genome is evaluated in all cells that are present in the system at that time. Cells are chosen for updating ordered by their age.

Development will take place in a finite planar environment. We call this two-dimensional environment the *development grid*. The development grid consist of square elements (*grid units*). A development grid with four by four square units is illustrated on the left hand side of Figure 2. Outside this grid, an *outside environment* is defined, that can be regarded as a single grid element with default values for all its properties, which never change during development.

A *cell* covers a number of grid elements on the development grid. At the beginning of development, a single cell covers the entire development grid. By cell divisions, this situation can later develop into multiple cells sharing the available area of the grid. Figure 2 shows the development grid, with a single cell occupying six grid elements. Each cell has an age, a number of proteins at a certain concentration, and a concentration of nuclear determinants, which are simplified to chemical substances that are present in gradient concentrations in the initial cell. By cell division, the gradient of the nuclear determinants is splitted into the child cells, providing them with a sense of which piece of the initial cell the descend from.

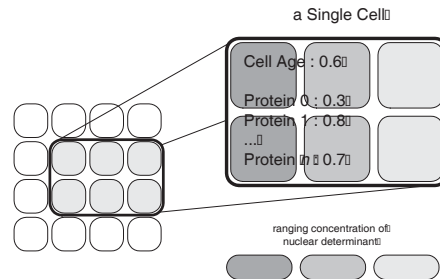


Figure 2. Detailed look at a single cell, with its cell age, the concentration of a number of proteins, and the ranging concentrations of a single nuclear determinant across the cell surface.

Genes in the model take the form of rules, with an antecedent and consequent. The antecedent implements the *controlling region* of the gene, while the consequent implements the *coding region* [5]. Antecedents reference any value in the development system, such as cell ages, protein concentrations, development time, etc. Gene consequents can result in changing protein concentrations, cell divisions, of differentiation of the CA transition rule inside the cell. A few examples of genes are should clarify the general concepts :

- (a) [`<age E> >= 0.6`] -->
change_prot:[nr:1 delta: +0.1]]
- (b) [`0.4 < <prot E:0>`] -->
divide:[orient:E-W,size:0.5, asymm,
effect:change_prot:[nr:0 delta: -0.2]]]
- (c) [`<global 0> < <dev_clk>`] -->
activate_TT_entry(14)

4 Modules in Development

Evolutionary computation operates based on the premise that there is a correlation between the fitness of a parent and its child that results after application of a genetic operator. For this to be possible, previous achievements in optimising parts of the genotype must be left standing, while others are subject to perturbation by the operator. Conceptually, this requires modularity in the genome.

We thus set out to investigate gene interaction in a setup of gene regulated development. Often in EC, gene interaction is discussed in the light of fitness contributions. In our setup of gene-regulation however, the interpretation of a gene is not a given fact, and gene interaction must first be discussed in terms of how genes influence each others activation itself. This brings us to the concept of activation propagation :

Definition 1 *Activation propagation is the effect by which activation of a source gene S causes the activation of target gene T during development.*

Based on this concept, we can regard two genes as interacting, when one's activation causes a propagation to activate the next. source gene propagates to a target gene, with which it interacts. This mechanism will be the atomic component by which gene interactions will be composed. To discuss this mechanism, we will limit ourselves to genes interaction through protein concentrations. Consider two genes S and T of the following form :

```
(S) ... --> change_prot:[ nr:0 delta: +0.3]]
(T) [0.3 >= <prot F:0>] --> ...
```

Our aim is to identify when and how an activation of S will cause a propagation to an activation of T . This type of interaction can be given a weight, indicating the degree to which activation propagation is likely to occur from the source gene to the target. Further details regarding the procedure to determine the interaction weight are discussed elsewhere [4]. This procedure

allows for identification of the interaction weights between all possible genes in the genome. Based on this, cascades of genes can be identified that, once started, have a high probability of begin expressed completely due to the effect of activation propagation. Cascades are regarded separate from each other, when the last activated gene in the cascade is unique to the specific cascade. This one, or more, unique genes at the end of the cascade can be regarded as the typical result of the cascade, while earlier genes in the cascade can be only leading to the this end result. These cascades are the most elementary modules present in gene-regulated development. Earlier genes in the cascades can be shared between cascades, as a result of the definition. It can be illustrated easily that perturbation of such shared genes causes their effects to propagate more extensively. This is illustrated in Figure 3, where the number of changes in activation as a result of a perturbation of the focal gene is plotted against the number of gene cascades the focal gene is shared in (we refer to [4] for more extensive discussion on these experiments).

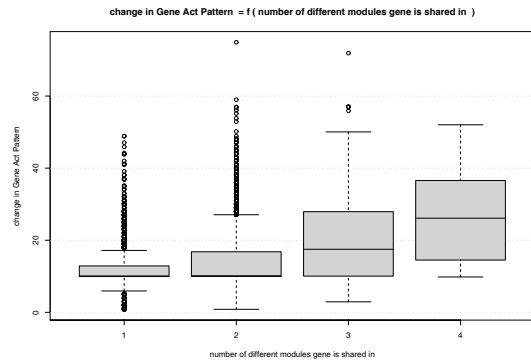


Figure 3. The change in the activation pattern as a function of the number of gene cascades the gene is shared in.

5 Modular Interaction, Development and EC

Having identified atomic modules in development as a representation, the next question is how these cascades interact with each other in the light of fitness of the phenotype. Watson discusses different types of module interaction and distinguishes *separable* modules on one end of the spectrum, and completely *non-decomposable* modules on the other. In between these extremes, modules are said to be *interdependent*, meaning that only a subset of all possible configurations of

module A is maximal in combination with the context provided by module B , and vice versa. Interesting for our purposes is that Watson shows that compositional mechanisms (such as crossover, given the right population structure) can allow easy evolution of solutions to problems that exhibit modular interdependence. Based on this, the discussion on gene-regulated development in combination with EC boils down seeing gene-regulated development genomes as consisting of modules with different types of interactions.

Intuitively it is easy to understand that gene cascades, if at all decomposable in relation to fitness, are likely to exhibit modular interdependence : the fact that that the outcome of one cascade is the input to the effect of the next severely diminishes the probability that the optimal configurations for different cascades are independent of each other.

To answer whether development can be decomposed into modules, and what type of interaction these would exhibit, we suggest two different ways to look at development in a top-down fashion :

1. Development can be seen as a mechanism that creates new cells, and gives them the correct identity. These two tasks can be seen separately. Without going into the details of this here, we indicate that it is possible, while not trivial, to condition genes in such a way that they adhere to one of these goals only and not both.
2. From a biological point of view, development is sometimes simplified as a process that dictates cell identity based on lineage and induction :
 - lineage being the determination of cell identity based on the cells line of descent
 - induction being the determination of cell identity based on local cell communication i.e. through the cell's environment

While both these views provide a possible inspiration of organising development in such a way that different modules have clearly separated and possible decomposable roles, none offers the advantages of answering the question by structural modularity of the phenotype. If we assume that the resulting phenotype is modular, different cascades can be organised to exert their influence in each of these modules, which is easier to organise than the previous suggestions, and is more scalable. In order to achieve structural modularity in the phenotype by means of development, mechanisms such as the following are indispensable :

- *a sense of position* : Nature provides implicit positional information[6], and so does our development

model by means of the gradients of nuclear determinants in the initial cell. When position can be referenced, it can be a source of limiting the activity of certain cascades to specific regions.

- *parallel channels of communication* : Genes communicate through proteins, for example using them to influence each others activation. In different phenotype modules, different proteins must be used to communicate, in order to avoid 'foreign' genes to be activated in the module. A multitude of possible proteins available for communication is thus necessary.
- *patterning cascades* : Specific cascades must lay out the global pattern of module boundaries, within which further differentiation will occur. Such mechanisms, albeit extremely complex, are also present in natural development during early development for laying out the body plan.

While some conditioning is still necessary, achieving modular phenotypes with decomposable paths of differentiation by different cascades seems achievable. Provided the transition from phenotype to fitness is non-disruptive, this approach seems to offer favourable scalability properties when using development.

6 From Phenotype to Fitness

Finally, the phenotype must be used to determine a fitness value. Important here is that the fitness value is in some way a decomposable combination of the scores that result from the different phenotype modules. In order for separability of gene cascades of different phenotype modules to be present, scores of different modules must again be separable or in other words independently contributing to the total fitness. If this is not the case, clean-cut separation of modules in the phenotype degrades to non-separability of modules in the final fitness, which is a simple case of the weakest link in the chain determining the strength of the chain. Thinking back of the complete setup with development in the evolutionary loop in Figure 1, modularity must be present throughout the sequence of development, phenotype and fitness in order to be of any use, as also indicated by others in recent work [1].

References

- [1] Lee Altenberg. Modularity in evolution: Some low-level questions. *Modularity: Understanding the Development and Evolution of Complex Natural Systems*, 2003.

- [2] F. David Fracchia. Integrating Lineage and Interaction for the Visualization of Cellular Structures. In Janice E. Cuny, Hartmut Ehrig, Gregor Engels, and Grzegorz Rozenberg, editors, *Proc. 5th Int. Workshop on Graph Grammars and their Application to Computer Science*, volume 1073 of *LNCS*, pages 521–535. Springer-Verlag, 1996.
- [3] Piet van Remortel, Tom Lenaerts, and Bernard Manderick. Lineage and Induction in the Development of Evolved Genotypes for Non-Uniform 2D CAs. In *proceedings of the 15th Australian Joint Conference on Artificial Intelligence 2002*, volume 2557 of *Lecture Notes in Computer Science (LNCS)*, Canberra, Australia, 2002. Springer Verlag.
- [4] Piet van Remortel, tom Lenaerts, and Bernard Manderick. Gene interaction and modularisation in a model for gene-regulated development. In *(under review)*, 2004.
- [5] L. Wolpert. *Principles of Development*. Oxford University Press, 1998.
- [6] Lewis Wolpert. Pattern formation in biological development. *Scientific American*, 239(4):154–164, 1978.