Information Sciences 424 (2018) 250–272

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Epigenetic algorithms: A New way of building GAs based on epigenetics

Daniel H. Stolfi*, Enrique Alba

Departamento de Lenguajes y Ciencias de la Computación, University of Malaga, Spain

ARTICLE INFO

Article history:
Received 21 June 2017
Revised 14 September 2017
Accepted 3 October 2017
Available online 5 October 2017

Keywords:
Evolutionary algorithm
Epigenetics
MKP
Bio-inspiration

ABSTRACT

This article presents a new set of ideas on how to build bio-inspired algorithms based on the new field of epigenetics. By analyzing this domain and extracting working computational ideas we want to offer a set of tools for the future creation of representations, operators, and search techniques that can competitively solve complex problems. To illustrate this, we describe an epiGenetic Algorithm, analyze its behavior and solve a set of instances of the multidimensional knapsack problem. Since we are in some measure opening a new line of research, we include a description of epigenetics and computational search, show their working principles and show an example algorithm solving a real problem. Our aim is to offer ideas as well as put them to work, to show that they are actually competitive, not just a nice new inspiration.

1. Introduction

In a letter to T.H. Huxley (Dec. 28th 1859), C. Darwin wrote that “Science is so narrow a field, it is clear there ought to be only one cock of the walk!”. He was referring to his previous competitor, R. Owen, and was obviously talking about himself as the new leader in biology after the recent publication of *The Origin of Species*. The basis for the (now traditional) conception of inheritance, genomes and the evolution of the fittest were then widely accepted, and swept aside other theories and beliefs like a huge hurricane.

Darwin, and many others before (e.g., Owen) and after him (e.g., Weismann), contributed to dismissing some of the ideas of de Lamark [41]. In fact, several of these ideas were clearly, correctly discredited, such as the linear concept of evolution (known as *recapitulation*). The inheritance of acquired characters from Lamarck was also dismissed at this moment until now [54]. *Epigenetics* (relating heritable traits that respond to the environment and cannot be explained by changes in DNA sequences) is here to stay, after so much evidence of its existence and so many applications in biology, pharmacology and medicine [23,25,53].

It seems that we *have to* consider a second “cock of the walk” for a better understanding of nature, and that is what we do in this article: we combine the ideas of Darwin and Lamarck into one single computational algorithm, obviously within the niche of research into nature-inspired algorithms. In addition to a nice, appealing inspiration for future work, we of course aim for a competitive evaluation of the resulting techniques, to prove that not only are epigenetic algorithms new as a class (not a renaming of an existing algorithm!), but also that they are useful for solving combinatorial and other kinds of problems in modern research.

---

* Corresponding author.
E-mail addresses: dhstolfi@lcc.uma.es (D.H. Stolfi), eat@lcc.uma.es (E. Alba).

https://doi.org/10.1016/j.ins.2017.10.005
0020-0255/

Metaheuristics for combinatorial optimization problems [12] are frequently inspired by natural processes such as Darwins' theory of evolution: *evolutionary algorithms* are today a classic example [6]. They are used to solve highly complex real-world problems. Usually, they start with a set of initial candidate solutions and iteratively generate new ones in a chain of increasingly fitted populations towards the optimum of the problem. Their non-deterministic guided and intelligent search balances the exploration of the search space and exploits its more promising regions, to hopefully find the optimal solution to the problem being solved.

As mentioned, evolutionary algorithms (EAs) are stochastic search methods which solve a wide range of combinatorial (and continuous) problems very efficiently and effectively, especially those that cannot be solved, using classic optimization tools: guided or efficient exhaustive enumeration like in A*, mathematical programming, branch-and-bound and dynamic programming, to name a few. The multiple search at the same time, the representation of the solutions in any convenient manner for the search, the absence of requirements of continuity/derivability of the function being optimized, and many other advantages (like dealing with any type of constraints, mixed variable domains, large dimensionality, etc.) make EAs a great tool in modern research.

Some algorithms work better with specific types of problems and perform worse over others [62]. Consequently, this first study is targeted to develop a new bio-inspired algorithm family based on epigenetics, that can be later adapted to different problems. This is possible not just in the traditional manner (as with EAs), but also by using a variety of epigenetic mechanisms that we successfully translate from biology to computer optimization.

In summary, this article studies the diverse epigenetic mechanisms controlled by specific DNA methylation as a method of modifying DNA expression, that may be reversible and inheritable. From this study, we define a methodology to generate EAs that efficiently solve many different problems by using epigenetic concepts on information representation, such as histones, nucleosomes, and chromatin, and epigenetic operations such as genomic imprinting, reprogramming, paramutation, position effect, X-inactivation, bookmarking, and gene silencing. As initially stated, we move from the pure inspiration to the traditional optimization analysis, to ensure that this idea is noteworthy to know and usable in the future of bio-inspired computing.

The rest of the paper is organized as follows. In the next section, we review the state of the art related to our work. In Section 3 we introduce the epigenetic theory and in Section 4 we enumerate the Epigenetic Operators. Section 5 presents the epiGenetic Algorithm and its operators. Section 6 contains the description of the Multidimensional Knapsack Problem and Section 7 describes the competitors algorithms. After that, we comment on the parameterization of the algorithms in Section 8, discuss the preliminary evaluation of our algorithm in Section 9, and finally, conclusions and future work are given in Section 10.

2. Background

Epigenetics have inspired several articles in computer science and bioinformatics in the last decade [1], we comment on some of them here. We will show that, even if the term has been mentioned in the past, nothing similar to a methodology for building efficient algorithms has been developed in these few papers.

In [45] the authors describe the optimization strategies that bio-molecules utilize and propose an intragenerational epigenetic algorithm based on them. The authors also present an agent-based cell modeling and simulation environment, called SwarmCell, whose model has been built as an autopoietic system that represents a minimal biological cell. Then, they implement the epigenetic strategies in the model to better clarify the disease mechanisms at the sub cellular level. This strategy's proposed use is to study cancer development patterns in different cell types which have been differentiated by various trans-generational epigenetic mechanisms. The authors state that their epigenetic algorithm can prove to be a fundamental extension to existing evolutionary systems and swarm intelligence models. They discuss improving problem-solving capabilities by implementing epigenetic strategies in their SwarmCell model. Finally, for future work, they intend to develop a trans-generational epigenetic algorithm to demonstrate how the internal organization of a system can pass on its traits to the next generation. Although epigenetic techniques are also proposed in their work, our study focuses on a new set of algorithms based on natural evolution rather than autopoietic systems and swarm models.

An epigenetic approach in artificial life (ALife) is presented in [57] where the model proposed (EpiAL) uses a dynamic environment to influence the regulation of organisms and the possible inheritance of epigenetic acquired marks. The objective of the EpiAL model is to study the plausibility for the existence of epigenetic phenomena and its relevance to an evolutionary system, from an ALife point of view. Therefore, each agent is able to modify its phenotypic expression due to environment conditions, pass on epigenetic marks between generations enabling the existence of acquired traits which can be transmitted through consecutive generations of agents. The experimentation performed with the EpiAL model in order to study the mechanisms that influence the evolution of the agents shows that the epigenetic populations are able to regulate themselves for dynamic conditions, while the non-epigenetic populations find it hard to prosper in dynamic environments. The authors plan a future development of the model with a focus on both, biological knowledge (developmental biology) and possible problem solving techniques (dynamic environments). This epigenetic approach is focused on the evolution of epigenetic agents to gain more knowledge about this field while we propose an algorithm and solve problems with it.

The authors in [59] incorporate an explicitly controlled gene expression through histone modification in strongly-typed genetic programming (STGP) and call it, epigenetic programming. They propose a double cell representation of the simulated individuals represented by their respective chromatin structures. The authors view their proposed approach of epigenetic