Fitness Value Curves Prediction in the Evolutionary Process of Genetic Algorithms

Renuá Meireles Almeida
Federal University of Pará
Tucuruí, Pará
renua.almeida@tucurui.ufpa.br

Denys Menfredy Ferreira Ribeiro
Federal University of Pará
Tucuruí, Pará
denys.ribeiro@tucurui.ufpa.br

Rodrigo Moraes Rodrigues
Federal University of Pará
Tucuruí, Pará
rodrigo.rodrigues@tucurui.ufpa.br

Otávio Noura Teixeira*
Federal University of Pará
Tucuruí, Pará
onoura@gmail.com

ABSTRACT

The evolutionary process of a Genetic Algorithm (GA) depends on several factors, the initialization parameters are some of them [4]. One way to inspect the evolutionary process of a GA is to analyze the maximum, average, and minimum fitness of each generation. This article focused on the use of a machine learning model to predict the maximum, average, and minimum fitness values during the evolutionary process of a GA, and this, only with the knowledge of its initialization parameters. In order to accomplish this goal, a Random Forest model was trained with data from different GA executions for a given problem. The prediction process was performed with a very promising performance, where the challenge of predicting the evolutionary process of a GA was fulfilled with low error rates. This approach opens up several opportunities for advances in the segment, and in a way, contributes to the investigation and improvement of GAs, as well as demonstrating the importance of monitoring and storing the information generated during their evolutionary process.

CCS CONCEPTS

• **Computing methodologies** → **Genetic algorithms**; *Machine learning*;

KEYWORDS

Genetic Algorithms, Machine learning, Fitness, Prediction

ACM Reference Format:

Renuá Meireles Almeida, Denys Menfredy Ferreira Ribeiro, Rodrigo Moraes Rodrigues, and Otávio Noura Teixeira. 2021. Fitness Value Curves Prediction in the Evolutionary Process of Genetic Algorithms. In *2021 Genetic and Evolutionary Computation Conference Companion (GECCO '21 Companion), July 10–14, 2021, Lille, France.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3449726.3459539

*Advisor.

GECCO '21 Companion, July 10–14, 2021, Lille, France

ACM ISBN 978-1-4503-8351-6/21/07.
https://doi.org/10.1145/3449726.3459539

1 INTRODUCTION

GAs, presented in [3], are stochastic search and optimization methods based on biological evolution and inspired by Darwin's theory of "survival of the fittest" [2], in which there is a set of individuals/solutions that undergo a series of operations.

For the execution of a simple GA [2] some basic parameters have to be defined, more specifically: population size, crossover probability, mutation probability and a stopping criteria, which among several possibilities, can be defined by a fixed number of generations. And, the performance of the algorithm has strong links to the definition of these elements [4, 5].

A simplified way to inspect the evolutionary process of a GA is to record the maximum, average and minimum fitness values of individuals in each generation. With that, it is possible to plot a fitness chart by generation.

Due to their stochasticity, GAs are usually executed several times under the same parameterization in order to infer more reliably about the behavior and results obtained [4]. The reuse of all the experience generated in the process does not occur, and consequently, works that use the same problems already used by other researchers, whether to make comparisons with other algorithms or implement new mechanisms that improve the evolutionary process [6], usually pass by the same exhaustive process of repeated executions, in order to obtain knowledge about the GA performance under certain parameters conditions.

Based on this, this work aims to reuse the experiences obtained in GA executions, and to train a machine learning model to predict the maximum, average and minimum fitness curves that emerge during the evolutionary process of a GA, and this, only with the knowledge of its initialization parameters.

Not far from this idea, as explored in [1], there are some approaches that commonly use fitness predictors through surrogate, approximation and meta-models as alternatives for the original fitness functions calculations, which consequently implies a series of problems to be handled.

2 METHODOLOGY

The current generation (Current Gen) and 7 initialization parameters have been defined as features for the prediction model. The chosen parameters were the population size (*PopSize*), stop generation (*StopGen*), crossover probability (*CxPb*), mutation probability

Table 1: Interval restrictions for initialization parameters

Feature	Minimum value	Maximum value
PopSize	40	70
StopGen	40	70
CxPb	50,0%	90,0%
MxPb	0,0%	40,0%
Elitism	0	20
LowerBound	-50,0	0,0
UpperBound	0,0	50,0

($MxPb$), elitism ($Elitism$), the lower ($LowerBound$) and upper bound ($UpperBound$) of the adopted fitness function.

Despite the existence of several observable metrics during the evolutionary process of GAs, the targets defined for prediction were defined as the maximum ($MaxFitness$), average ($AvgFitness$) and minimum ($MinFitness$) fitness values of a certain generation. This, due to the facility of quick visualization of the convergence processes of GA.

Due to the absence of a specific data set for this solution, it was necessary to build it, and this was done through the collection of data generated in the execution of several random instances for a GA. Altogether, 5.000 executions were defined for the maximization of a real function given by $f(x) = (x^2 + x) \cdot cos(x)$, and in each, the initialization parameters were uniformly randomized under the restriction of predetermined intervals, these definitions can be seen in table 1. In each run, G records are added to the data set, where $G = StopGen$, totaling around 300.000 records were obtained.

A Random Forest model was trained with a random sampling of 85% of the data set and validated with the remaining 15%. The metrics used to measure this initial performance were the Coefficient of Determination (r^2) and the Mean Absolute Error (MAE).

In order to perform the prediction of the entire evolutionary process of a GA, it was necessary to generate the features that feed the model for each generation having only the initial execution parameters. To do this, simply replicate the initialization parameters for all generations, and for feature $CurrentGen$ a sequence of integers from 1 to G is generated, where $G = StopGen$.

To carry out this test, a random parameterization for GA was generated, maintaining the same rules imposed in the generation of data for training and model validation, and it was given by 45, 65, 0, 82, 0, 16, 9, $-9, 0$ and $3, 0$ for $PopSize$, $StopGen$, $CxPb$, $MxPb$, $Elitism$, $LowerBound$ and $UpperBound$ respectively.

3 RESULTS

The model's performance is shown in Table 2. It is noticed that the target $MaxFitness$ is the value that the model had an easier time predicting when compared to other features, this is mainly due to the regularity of the curve generated in the evolution process, unlike what happens with $MinFitness$, which has a highly irregular, presenting abrupt changes, which consequently makes it the target with the highest error rates. These behaviors can be better visualized in figure 1, which presents the results of the prediction of the entire evolutionary process for the proposed parameter setting, where the model obtained a MAE of $6, 24$.

Table 2: Random Forest validation performance

Metric	MaxFitness	AvgFitness	MinFitness	Mean
r^2	0,991	0,965	0,710	0,889
MAE	16,20	58,33	218,35	97,63

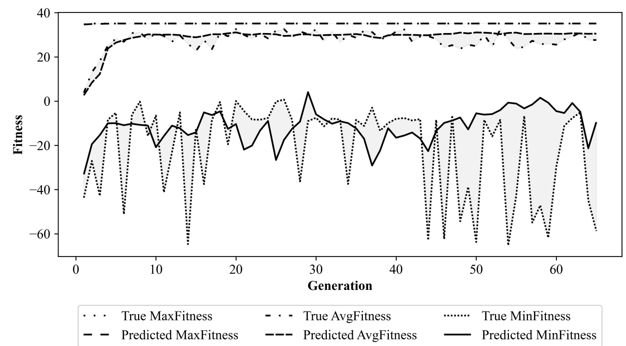

Figure 1: Results of model predictions under parameter setting 1.

4 CONCLUSIONS AND FUTURE WORKS

The results presented so far (including many that cannot be shown here) indicate a great potential in this approach. That contributes with experimentation and research activities related to these algorithms. It also highlight the importance of monitoring, generating and storing the data that arise from the execution of these algorithms.

Despite the errors present in the predictions, mainly in the prediction of the minimum fitness values, in terms of graphical visualization, the predictions adopt behaviors very close to the real ones. This element opens possibilities for dynamic and immediate visualization of the influence of different initialization attributes.

There are good alternatives to be explored and deepened from this work, for example the analysis in the relation of quantity of data necessary for the model training, make use of more complex problems, among others.

REFERENCES

[1] Maumita Bhattacharya. 2013. Evolutionary Approaches to Expensive Optimisation. *International Journal of Advanced Research in Artificial Intelligence* 2, 3 (2013). https://doi.org/10.14569/IJARAI.2013.020308

[2] David E Goldberg. 1989. Genetic algorithms in search. *Optimization, and Machine-Learning* (1989).

[3] John Henry Holland et al. 1992. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence.* MIT press.

[4] Melanie Mitchell. 1998. *An Introduction to Genetic Algorithms.* MIT Press, Cambridge, MA, USA.

[5] Mohsen Mosayebi and Manbir Sodhi. 2020. Tuning Genetic Algorithm Parameters Using Design of Experiments. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion (GECCO '20).* Association for Computing Machinery, New York, NY, USA, 1937–1944. https://doi.org/10.1145/3377929.3398136

[6] Rodrigo Pereira, Daniel Souza, Marco Mollinetti, Mário Serra Neto, Edson Yasojima, Otávio Noura Teixeira, and Roberto Oliveira. 2020. Game Theory and Social Interaction for Selection and Crossover Pressure Control in Genetic Algorithms: An Empirical Analysis to Real-Valued Constrained Optimization. *IEEE Access* PP (08 2020), 1–1. https://doi.org/10.1109/ACCESS.2020.3014577