

Observing the Population Dynamics in GE by means of the Intrinsic Dimension

Eric Medvet¹, Alberto Bartoli¹, Alessio Ansuini², and Fabiano Tarlao¹

¹ Department of Engineering and Architecture, University of Trieste, Trieste, Italy

² International School for Advanced Studies (SISSA/ISAS), Trieste 34136, Italy

Abstract. We explore the use of Intrinsic Dimension (ID) for gaining insights in how populations evolve in Evolutionary Algorithms. ID measures the minimum number of dimensions needed to accurately describe a dataset and its estimators are being used more and more in Machine Learning to cope with large datasets. We postulate that ID can provide information about population which is complimentary w.r.t. what (a simple measure of) diversity tells. We experimented with the application of ID to populations evolved with a recent variant of Grammatical Evolution. The preliminary results suggest that diversity and ID constitute two different points of view on the population dynamics.

Keywords: Diversity, Fitness Landscape, Initialization, Grammatical Evolution

1 Introduction and related work

In Evolutionary Computation (EC) a population of candidate solutions to a problem, also known as *individuals*, evolves by mutating and recombining fitter individuals. It is well known among EC scholars and practitioners that if the population is composed of many very similar individuals, i.e., if it is not *diverse*, the exploration phase of the evolution may be ineffective, eventually leading to the so called *premature convergence*. Lack of diversity has been acknowledged as a problem of Evolutionary Algorithms (EAs) since the beginning of the EC research field and has been widely studied [1]. In particular, different ways have been proposed for measuring the population *diversity*.

In this work, we explore the use of the Intrinsic Dimension (ID) for characterizing the structure of the population during an EA run and understanding if and how it relates with diversity. ID can be seen as the minimum number of variables which are needed to accurately describe the features of a system [2] and has been used in different domains in which data is, in general, large. Often, ID has been proposed as a mean for reducing the dimensionality of the data with the aim of fighting the so called “curse of dimensionality” [3].

In the EC context, we postulate that measuring the ID of the population may help the understanding of the structure of the population in a more detailed way than measuring its diversity. The intuition is that the action of the selective pressure may impose a “shape” to the population during the evolution and that

that shape corresponds to a region of the search space of which dimension is much smaller than the full space dimension. From another point of view, during the evolution the individuals may move along one or more paths within the *fitness landscape*.

In order to gain insights about our speculation, we performed a set of experiments in which we considered a popular EA, Grammatical Evolution (GE) [4], and measured how the ID, the diversity, and the fitness of the best individual vary over the evolution. We considered GE for three reasons: (a) its suitability for any problem which can be described by a Context Free Grammar makes it an EA of great practical relevance; (b) it employs a bit string-based representation, hence the search space allows using a simple distance among genotypes (i.e., the Hamming distance between bit strings); (c) it has been showed to suffer from lack of diversity [5] and some methods for preserving it has been proposed too [6].

From our preliminary experiments, we found that ID seems to provide a complementary information w.r.t. diversity, the latter measured as the fraction of unique genotypes in the population. In particular, ID always increases during the evolution, whereas the diversity initially quickly decreases and later slowly increases when the problem remains unsolved. This finding suggests that the high diversity in the initial population might be of “poor quality”, since it does not correspond to an high ID—a finding which confirms the relevance of the population initialization in GE [7].

As the lack of diversity is perceived as a key problem with practical effects in the EC community, many studies have been devoted to it. In particular, different ways for measuring the diversity have been proposed—for space constraints, we refer the reader to [1] for a comprehensive analysis. To the best of our knowledge, no previous works exist on the application of the ID to a population of an EA.

2 Background: estimated Intrinsic Dimension

The ID of a dataset, i.e., a set of points defined in a high-dimensional space, can be seen as the minimum number of variables which are needed to accurately describe that dataset. Measuring the actual ID of a real dataset may be an hard task, due to the presence of noise.

In order to deal with noise and to reduce the computational burden of computing the actual ID, many ways of estimating it have been proposed. We here use TWO-NN, a recent estimation method which requires to compute only the distance of each point to its two nearest neighbors (NNs) [2].

In detail, TWO-NN works as follows. Let P be the set of n items whose ID has to be computed and let D be the corresponding distance matrix, i.e., $d_{i,j}$ is the distance between the i th item $p_i \in P$ and the j th item $p_j \in P$. First, the ratios $\mathbf{r} = (r_1, \dots, r_n)$ between the distances of the two closest items for each item is computed as $r_i = \min_{j \neq i, h \neq i} \frac{d_h}{d_j}$, with $d_h \geq d_j$. Second, \mathbf{r} is sorted in ascending order. Third, the values of $\mathbf{x} = \log \mathbf{r}$ and $\mathbf{y} = -\log(1 - \frac{1}{n}\{0, 1, \dots, n-1\})$ are computed. Fourth and finally, the points defined by \mathbf{x}, \mathbf{y} are fitted with a straight line and the resulting slope is taken as the ID of the points in P .

In our work, P is composed by a population of bit strings observed during an EA run. This exacerbates the estimation of the ID for two reasons: (i) the Hamming distance among individuals is discrete, rather than continuous; (ii) P may contain many duplicated individuals, due to low diversity in later stages of the evolution. As a result, the quality of the linear fitting of the \mathbf{x}, \mathbf{y} points, expressed in terms of the product-moment coefficient of correlation, can be low. In order to mitigate the latter cause, we computed the ID on the set of unique genotypes, rather than on the entire population.

3 Experiments, results, and discussion

We experimented with a recent variant of GE, called Weighted Hierarchical GE (WHGE) [8] which differs from the original only in the representation internals and has been shown to perform better. We considered two benchmark problems with tunable hardness: the Parity problem, for which we experimented with the number of input bits in $b \in \{3, \dots, 7\}$, and the KLandscape synthetic benchmark [9], for which we set $k \in \{3, \dots, 7\}$. For both, the lower the parameter, the easier the problem.

For each problem, we executed 5 independent runs and measured, at each generation, the fitness of the best individual, the population diversity (i.e., the fraction of unique genotypes), and the population ID. We used the standard $m + n$ replacement strategy with $m = n = 500$ individuals, which we evolved for 50 generations, and a tournament selection with size 3. We set the genotype size to 256 bit. Figure 1 shows the results for the two problems, KLandscape (above) and Parity (below)—for the best fitness, the lower, the better.

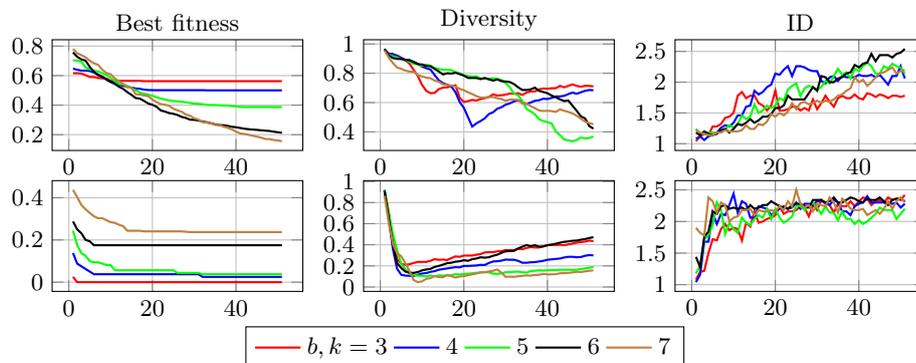


Fig. 1. Results, averaged across the 5 runs, for KLandscape (above) and Parity (below).

It can be seen the way ID and diversity vary during the evolution is different. The diversity starts at ≈ 1 in the initial population and then decreases: for Parity, the decrease is very rapid, whereas it is slower for KLandscape; in both cases, the

diversity reaches a minimum and inverts the trend when the best fitness ceases to (significantly) improve. Instead, ID starts at ≈ 1 and steadily increases: when diversity decreases fast, ID increases rapidly; otherwise it increases slowly.

We believe that the diversity starts increasing when the evolution stagnates, i.e., when the best fitness stops improving, because of the representation degeneracy (the tendency of mapping different genotypes to the same phenotype [5]): many different genotypes mapping to the same best phenotype start to fill the population. Interestingly, ID seems not to be affected by this phenomenon, as it remains roughly steady in later stages of the stagnating evolutions.

Moreover, final values of ID are always greater than initial values. No sharp conclusions can be drawn in these respects, but we believe that ID values suggest that the final population tend to lie on flat regions in the search space. On the other hand, the fact that the highly diverse initial populations exhibit a low ID suggests that the population initialization is indeed a critical component of GE, as already observed [7].

References

1. Squillero, G., Tonda, A.: Divergence of character and premature convergence: A survey of methodologies for promoting diversity in evolutionary optimization. *Information Sciences* **329** (2016) 782–799
2. Facco, E., d’Errico, M., Rodriguez, A., Laio, A.: Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Scientific reports* **7**(1) (2017) 12140
3. Campadelli, P., Casiraghi, E., Ceruti, C., Rozza, A.: Intrinsic dimension estimation: Relevant techniques and a benchmark framework. *Mathematical Problems in Engineering* **2015** (2015)
4. Ryan, C., Collins, J., O’Neill, M.: Grammatical evolution: Evolving programs for an arbitrary language. In: *European Conference on Genetic Programming*, Springer (1998) 83–96
5. Medvet, E.: A comparative analysis of dynamic locality and redundancy in grammatical evolution. In: *European Conference on Genetic Programming*, Springer (2017) 326–342
6. Medvet, E., Bartoli, A., Squillero, G.: An effective diversity promotion mechanism in grammatical evolution. In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, ACM (2017) 247–248
7. Nicolau, M.: Understanding grammatical evolution: initialisation. *Genetic Programming and Evolvable Machines* **18**(4) (2017) 467–507
8. Medvet, E.: Hierarchical grammatical evolution. In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. GECCO ’17, New York, NY, USA, ACM (2017) 249–250
9. Vanneschi, L., Castelli, M., Manzoni, L.: The k landscapes: a tunably difficult benchmark for genetic programming. In: *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, ACM (2011) 1467–1474