Visual Analytics of EA Data

Evelyne Lutton
AVIZ Team, INRIA Saclay - Île-de-France,
Bat 490, Université Paris-Sud
F91405 ORSAY Cedex, France
Evelyne.Lutton@inria.fr

Jean-Daniel Fekete
AVIZ Team, INRIA Saclay - Île-de-France,
Bat 490, Université Paris-Sud
F91405 ORSAY Cedex, France
Jean-Daniel.Fekete@inria.fr

ABSTRACT

An experimental analysis of evolutionary algorithms usually generates a huge amount of multidimensional data, including numeric and symbolic data. It is difficult to efficiently navigate in such a set of data, for instance to be able to tune the parameters or evaluate the efficiency of some operators. Usual features of existing EA visualisation systems consist in visualising time- or generation-dependent curves (fitness, diversity, or other statistics). When dealing with genomic information, the task becomes even more difficult, as a convenient visualisation strongly depends on the considered fitness landscape. In this latter case the raw data are usually sets of successive populations of points of a complex multidimensional space. The purpose of this paper is to evaluate the potential interest of a recent visual analytics tool for navigating in complex sets of EA data, and to sketch future developements of this tool, in order to better adapt it to the needs of EA experimental analysis.

Categories and Subject Descriptors

I.2 [**Problem Solving, Control Methods, and Search**]: Heuristic methods; H.5 [**User Interfaces**]: Graphical user interfaces (GUI)

General Terms

Experimentation

Keywords

Visual Analytics, experimental analysis of EAs, parameter tuning, fitness landscape visualisation.

1. INTRODUCTION

The crucial problem of parameter tuning and the huge amount of information that can be collected during the run of an Evolution Algorithm (EA) has led authors to propose various visualisation software. When reviewing this literature or experimenting existing software and toolboxes [1, 2, 3, 4], a series of important issues seems to remain difficult:

- How to visualise individuals, populations (large search spaces, combination of symbolic/numeric, discrete/continuous values). Proposed solution are usually problem dependent (signals, images, sounds, graphs, networks, etc ...) [5].

- How to visualise history, evolutionary mechanisms, and transmission of genetic material [6, 7].

- How to visualise non-standard algorithms : multi-objective EAs, coevolution algorithms, ACO, ...

In this short paper, we investigate the use of a recent tool, ScatterDice/GraphDice, developed by the visual analytics community, that may provide efficient and generic answers to some of the previous challenges. An analysis of population clouds is briefly described in section 2 for some classical test-functions. Tests are based on the EASEA language [8][1]. Section 3 then gives future work we intend to develop on this topic.

2. ANALYSING POPULATIONS CLOUDS USING SCATTERDICE/GRAPHDICE

ScatterDice [9] is a multidimensional visual exploration tool, that allows the user to navigate in a multidimensional set via simple 2D projections, organised as scatterplot matrices, visual coherence between projections is based on animated 3D transitions. Various queries can be built using bounding volumes, and refined from different views[2]. A recent evolution of ScatterDice using the same principles but with many additional features is GraphDice [10]. It allows reading the same type of data (.csv files), and other more sophisticated formats, as it also embeds graph visualisation utilities[3].

This tool has been used to visualise data collected during the run of an EA. At each generation, the composition of the current population has been saved in a simple ".csv", forming a "cloud" of successive populations. The test functions used in the experiments are Weierstrass functions in dimension 2, with Hölder exponents $H = 0.2$ (very irregular) and $H = 0.9$ (more regular), and a Rosenbrock function in dimension 10. The genetic engine is a simple generational algorithm on R^n using tournament selection, geometric (barycentric) crossover and log-normal self-adaptive mutation. Additional dimensions are thus considered in the search space, the σ_i values, that represent the mutation radius for each coordinate x_i. The population size is 100 and the algorithm runs for 100 generations. This genetic engine is available in the sample programs (weierstrass.ez) distributed with the EASEA software.

A visualisation of the population cloud for the 2D Weierstrass function of dimension 0.2 is given on figure 1. The scatterplot matrix on the left of the figure, gives an overview of the possible visualisations. The columns and lines of this matrix can be dragged and dropped as wished by the user, a default order is proposed, based on an algorithm that reduces "clutter" [10]. On the right, a detailed view is given, on which some queries have been visualised: in red, the points corresponding to the 10 first generations of the run, in yellow, the 10 last generations, and in green, the best fitness

---

[1] http://sourceforge.net/projects/easea/
[2] http://www.aviz.fr/~fekete/scatterdice/
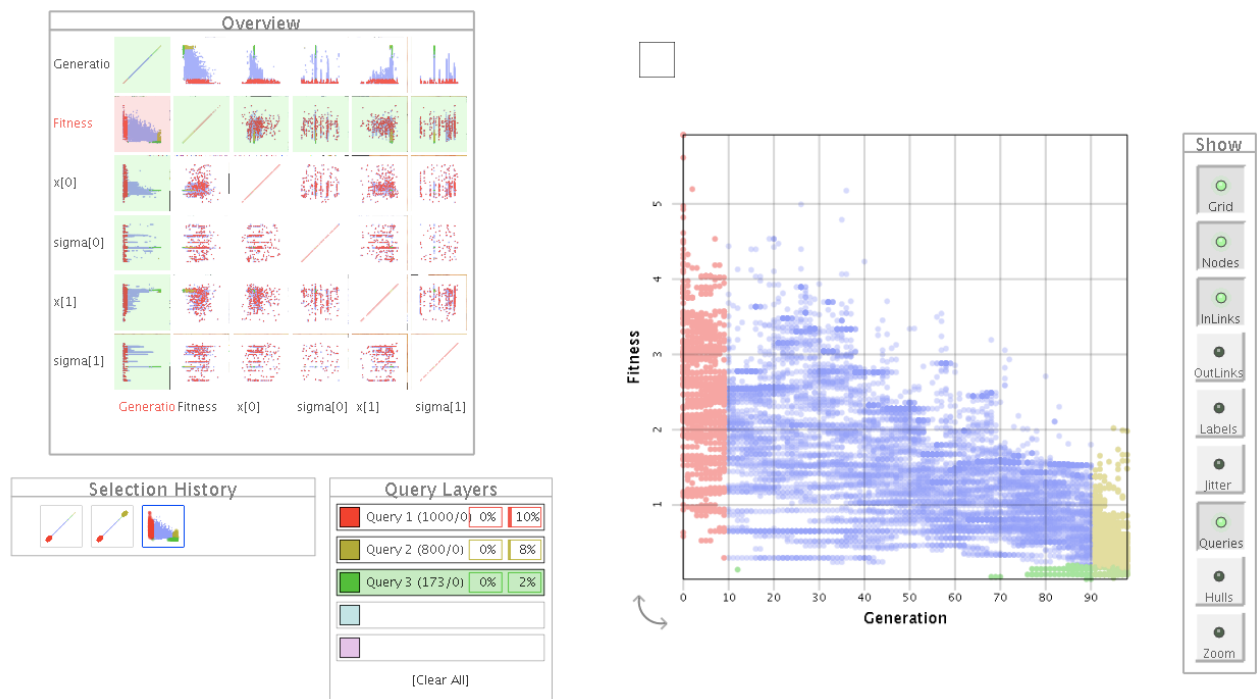[3] http://www.aviz.fr/graphdice/

Figure 1: 2D Weierstrass function of Hölder exponent 0.2. Scatterplot and Fitness versus generation view. Red points correspond to the first 10 generations, yellow points, to the 10 last ones, and green points, to the best fitness areas.

points. The queries are organised as layers, the "Query layers" window gives the details of the three queries, with some additional measurements (percentage of selected points, and percentage of selected edges if a graph is visualised : GraphDice actually considers a set of points as a degenerate graph, made of a collection of nodes with no edges). Bottom left, a "Selection History" window show how the queries have been sculpted: queries 1 and 2 have been activated on the Generation versus Generation view, i.e. the top left plot of the scatterplot matrix, while query 3 has been made on the fitness versus generation plot. On the extreme rigth of the window, a toolbar proposes various visualisation options (the "Show" window), for instance "Grid" activates a grid on the dataset, "Labels" allows to display attributes attached to a point. It is thus for instance possible to identify different runs of an EA using a label, and use this option to separate the data when needed. "Hull" displays a convex hull for each query, and "Zoom" activates an automatic zoom focussed on the selected data.

3. CONCLUSION AND FUTURE WORK

We have shown above that GraphDice can easily be used for visual inspection of EA behaviour, and allows for instance a rapid visual inspection of exploration capability and population diversity. First experiments have been performed on toy problems, but let us identify a set of desirable features to adapt GraphDice to the specific needs of EA analysis. The following issues will guide future developments of a GraphDice version adapted to EAs:

- Tests have been performed on relatively small data sets (up to 100x100 individuals in 10 dimensional space). The scalability issue will be tested more extensively.

- Various usual statistics (per generation, per fitness level) and query-based statistics will be implemented, including comparison of distributions (p-values).

- GraphDice, that also allows visualising graphs, will be evaluated for the genealogy visualisation task.

- On-line visualisation issues will also be considered.

4. REFERENCES

[1] Pohlheim, H.: Visualization of evolutionary algorithms - set of standard techniques and multidimensional visualization. In: GECCO'99 - Proceedings of the Genetic and Evolutionary Computation Conference, San Francisco, CA: Morgan Kaufmann. (1999) 533–540

[2] Pohlheim, H.: Geatbx - genetic and evolutionary algorithm toolbox for matlab http://www.geatbx.com/.

[3] Collins, T.D. In: Visualizing evolutionary computation. Springer-Verlag New York, Inc., New York, NY, USA (2003) 95–116

[4] Spears, W.M.: An overview of multidimensional visualization techniques. In: Evolutionary Computation Visualization Workshop. (1999) T. D. Collins, editor, Orlando, Florida, USA.

[5] Mach, M. Zetakova, Z.: Visualising genetic algorithms: A way through the Labyrinth of search space. In: Intelligent Technologies - Theory and Applications. IOS Press, Amsterdam (2002) 279–285 P. Sincak - J. Vascak - V. Kvasnicka - J. Pospichal (eds.).

[6] Bullock, S., Bedau, M.A.: Exploring the dynamics of adaptation with evolutionary activity plots. Artif. Life **12** (March 2006) 193–197

[7] Hart, E., Ross, P.: Gavel - a new tool for genetic algorithm visualization. IEEE Trans. Evolutionary Computation **5**(4) (2001) 335–348

[8] Collet, P., Lutton, E., Schoenauer, M., Louchet, J.: Take it EASEA. In Schoenauer, M., Deb, K., Rudolf, G., Yao, X., Lutton, E., J.J., M., Schwefel, H.P., eds.: Parallel Problem Solving from Nature - PPSN VI 6th International Conference, Paris, France, Springer Verlag (September 16-20 2000) LNCS 1917.

[9] Elmqvist, N., Dragicevic, P., Fekete, J.D.: Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2008) **14**(6) (2008) 1141–1148

[10] Bezerianos, A., Chevalier, F., Dragicevic, P., Elmqvist, N., Fekete, J.D.: Graphdice: A system for exploring multivariate social networks. Computer Graphics Forum (Proc. EuroVis 2010) **29**(3) (2010) 863–872