An Overview of Multidimensional Visualization Techniques

William M. Spears
AI Center - Code 5514
Naval Research Laboratory
Washington, DC 20375

Abstract

Although visualizing evolutionary algorithms often provides very specific and unique challenges, it is not necessary to always reinvent visualization techniques from scratch. Instead we can often look to existing general visualization techniques for help in creating our own specific visualization tools. This paper provides a brief overview of several existing techniques for visualization.

1 INTRODUCTION

Visualizing evolutionary algorithms (EAs) provides very specific and unique challenges, due to the enormous amount of information that needs to be examined, and due to the high multidimensionality of the information. Also, often that information is discrete (for example, in EAs that use discrete representations), presenting further difficulties. However, rather than reinvent the proverbial visualization wheel, we can often look to existing general visualization techniques for help in creating our own specific visualization tools.

Visualization does not operate in a vacuum. One hand, theoretical and empirical considerations drive what we want to visualize. On the other hand visualization can provide new insights, which influences the theories we want to develop and the experiments we want to perform. The purpose of this short abstract is to provide a quick overview of several existing techniques for visualization. A much more thorough overview can be found in Wegman and Carr (1992).

2 COLOR

Color is a very useful mechanism for adding a dimension to almost any visualization technique. Care must be taken if the added dimension is ordered, since the color palette has no natural order. In these situations, gray scale or saturation provides the necessary ordering (Tufte, 1990).

One use of color is to paint each individual of an EA at generation 0 a unique color. Cloning creates an individual with the same color and recombination creates individuals with multiple colors (if the two parents have different colors). Mutation paints an allele white (no ancestral individual is colored white). Thus, this coloring mechanism allows the user to monitor the survival of original ancestral material. Interestingly, different recombination operators lead to different losses in ancestral material. There is no theoretical explanation for these results, so this provides a nice example of how visualization can suggest a need for new theories.

3 GLYPHS

One useful technique for displaying multidimensional data is through the use of "glyphs". Glyphs are small icons that change appearance according to the data they represent. A glyph is drawn for each multidimensional data point. Commonality and differences in the features of the glyphs can help illustrate important structures in the original data. Fienberg (1979) provides an overview of different forms of glyphs. Two popular forms are "star plots" and "Chernoff faces".

Star plots (also called "sunflower plots") are points with spokes emanating from the points. If there are n dimensions then there are n spokes, and the length of each spoke represents the magnitude of one variable.

A more complex example of glyphs is called a Chernoff face, in which each data point is represented with a stylized drawing of a human face. The size, shape, and separation between parts of the face (e.g., the size and shape of the eyes or their separation) represent the magnitude of the different variables. The motivation for Chernoff faces stems from the fact that humans are excellent at differentiating between and recognizing human faces. One nice extension of this work uses "Chernoff bodies" (Bob Daley, personal communication) to describe combinations of behavioral features of coevolved agent strategies. With Chernoff bodies the icons have bodies with arms and legs, in addition to faces. Because the application involved agents, features in the Chernoff bodies map semantically to obvious behavioral characteristics of the agents. For example, skinnier bodies meant that the agents were more efficient about getting to their goal.

4 PROJECTION TECHNIQUES

Another useful technique for displaying multidimensional data is to project that data into a much smaller subspace (generally of one, two, or three dimensions), in the hope that various multidimensional structures will reveal themselves in lower dimensions.

One projection technique is called "Andrews curves" (Andrews, 1972), in which each data point is mapped to a curve in 2-D space. This technique is useful because multidimensional points that cluster together will tend to have curves that lie close together.

Another projection technique simply projects the multidimensional data into a 2-D plane (defined by two orthogonal vectors). Smoothly rotating the 2-D plane (equivalently, rotating the data) causes the data points to smoothly move in the 2-D plane, often revealing unusual structures within the multidimensional data. "Grand tour" (Asimov, 1985) techniques attempt to traverse the space of all rotations, with little overlap.

Unfortunately, grand tour techniques are time consuming, since the space of all rotations is extremely large. An alternative is to automate the procedure for "interesting" projections of the data. This is referred to as "projection pursuit" (Crawford and Fall, 1990). Projection pursuit makes use of a fitness metric that describes the interestingness of the current projection, and then hill-climbs the space of projections, looking for locally optimal projections. I would propose that this would be an interesting application for an EA, especially a speciating EA (e.g., Spears, 1994) that could simultaneously find numerous good projections.

5 PARALLEL COORDINATES

The rationale for projection techniques is that it is impossible to view multidimensional data when the coordinate axes of the space are orthogonal. However, an alternative solution is to use parallel coordinate axes (Wegman, 1990). Each multidimensional point $(x_1, ..., x_n)$ is drawn by plotting x_1 on axis 1, x_2 on axis 2, etc., creating a broken line representation for each point. One nice consequence is that the transformation of points from Cartesian coordinates to parallel coordinates is a highly structured mathematical transformation, and hence mathematical objects are mapped to mathematical objects. For example rotations in Cartesian space become translations in parallel coordinate space.

An extension of this technique for time varying data is to animate a parallel coordinate plot, redrawing the multidimensional points as they change over time. This has been especially useful in our investigations of how the "column mass" of a Markov chain of an EA changes over time (Spears and De Jong, 1996).

References

D. F. Andrews (1972). Plots of high dimensional data. *Biometrics* **28**, 125–136.

D. Asimov (1985). The grand tour: A tool for viewing multidimensional data. *SIAM Journal Sci. Statistical Computing.* **6, 1**, 128–143.

H. Chernoff (1973). Using faces to represent points in k-dimensional space. *Amer. Statist. Assoc.* **68**, 361–368.

S. Crawford and T. Fall (1990). Projection pursuit techniques for visualizing high-dimensional data sets. In G. M. Nielson, B. Shriver and L. J. Rosenblum (eds.) *Visualization in Scientific Computing* IEEE Computer Society Press, 94–108.

S. Fienberg (1979). Graphical methods in statistics. *Amer. Statist.* **33**, 165–178.

W. Spears and K. De Jong (1996). Analyzing GAs using Markov models with semantically ordered and lumped states. *Foundations of Genetic Algorithms* **4**, 85–100.

W. Spears (1994). Simple subpopulation schemes. *Evolutionary Programming Conference*, 296–307.

E. Tufte (1990). Envisioning information. *Graphics Press.*

E. Wegman and D. Carr (1992). Statistical graphics and visualization. *Center for Computational Statistics, Technical Report #84, George Mason University.*

E. Wegman (1990). Hyperdimensional data analysis using parallel coordinates. *Amer. Stat. Assoc.* **85**, 664–675.