A Study of An Evolutionary Algorithm with Applications to Statistics

Sina Sanei, Stat 540 Project Proposal, Fall 2018

Evolutionary algorithms are part of a broader group of optimization techniques known as *Stochastic Optimization*.Stochastic optimization plays a significant role in the analysis, design, and operation of modern systems. Methods for stochastic optimization provide a means of dealing with inherent system noise and coping with models or systems that are highly nonlinear, high dimensional, or otherwise inappropriate for classical deterministic methods of optimization. Stochastic optimization algorithms have broad application to problems in statistics (e.g., design of experiments and response surface modeling), science, engineering, and business. Algorithms that employ some form of stochastic optimization have become widely available. For example, many modern data mining packages include methods such as simulated annealing and genetic algorithms as tools for extracting patterns in data.[1]

Evolutionary algorithms are used to find maxima of functions. Their strength is an ability to find a global maximum in the presence of local maxima. The computations do not require derivatives or convexity, but still may be fairly computationally intensive in larger dimensions. This article presents a new type of evolutionary algorithm that works well in many dimensions, with the added advantage that linear equality constraints are implemented in a natural way. [2] Bounds on the coordinates of the solution are also easy to implement.
EAs have a number of components, procedures or operators that must be specified in order to define a particular EA. The most important components are:

- representation (definition of individuals)
- evaluation function (or fitness function)
- population
- parent selection mechanism
- variation operators, recombination and mutation
- survivor selection mechanism (replacement)

Each of these components must be specified in order to define a particular EA. Furthermore, to obtain a running algorithm the initialisation procedure and a termination condition must be also defined.

I will start this project by implementing genetic algorithm on a toy problem similar to the one outlined in [3]. This will give me an idea how to keep track of the components of algorithm. Next, as a main part of my project, I will try to mimic a new evolutionary algorithm using linear constraints as introduced in [2]. In this Algorithm ,both recombination and mutation use the Cauchy distribution. The algorithm is implemented to solve four different problems, two of which are purely statistical problems, namely 1-Robust Regression , 2-Density Estimation. The author has used FORTRAN for implementations, and I will try to solve these problems with R software.

References

[1] Gentle, J. E., Hrdle, W. K., Mori, Y. (Eds.). (2012). Handbook of computational statistics: concepts and methods. Springer Science Business Media.

[2] Meyer, M. C. (2003). An evolutionary algorithm with applications to statistics. Journal of Computational and Graphical Statistics, 12(2), 265-281.

[3] Chatterjee, S., Laudato, M., Lynch, L. A. (1996). Genetic algorithms and their statistical applications: an introduction. Computational Statistics Data Analysis, 22(6), 633-651.