GENETIC ALGORITHM

By
Ben Shedlofsky

Agenda

- History of Genetic Algorithm
- Methodology of Genetic Algorithm
- Process of Genetic Algorithm
- Pros And Cons of Genetic Algorithm
- About Genetic Algorithm
- Going over an example and pseudo-code of the algorithm

History

- Computer simulations of evolution started in 1954
  - Work of Nils Aall Barriclelli
- 1957 Australian geneticist Alex Fraser published a series of papers
  - Simulation of artificial selection of organisms
    - Multiple loci control a measurable trait
- Computer simulation of evolution by biologist in the early 1960s

http://en.wikipedia.org/wiki/Genetic_Algorithm

History Continued

- The methods described in books by Fraser, Burnell and Crosby
- Fraser's simulations included all the essential elements of modern genetic algorithms
- Hans Bremermann published series of papers in the 1960s
  - Also adopted population of solution to optimization problems, undergoing recombination, mutation, and selection
    - Also included elements of modern genetic algorithms

http://en.wikipedia.org/wiki/Genetic_Algorithm

History Continued II

- Artificial evolution became widely recognized optimization method as result work
  - Ingo Rechenberg and Hans-Paul Schwefel
    - 1960s and early 1970s
    - Solve complex engineering problems through evolution strategies
- Genetic algorithms became popular through work
  - John Holland early 1970s

http://en.wikipedia.org/wiki/Genetic_Algorithm

History Continued III

- John Holland book Adaption in Natural and Artificial Systems (1975)
  - Work originated with studies cellular automata
  - Conducted by Holland and his students
  - Introduced formalized framework
    - Predicting quality of next generation
      - Known as Holland's Schema Theorem

http://en.wikipedia.org/wiki/Genetic_Algorithm

History Continued IV

- Research GA's remained largely theoretical
  - Until mid 1980s
    - The First International Conference on Genetic Algorithms was held
- Academic interest grew and increase desktop computational power allowed practical application new technique
- Late 1980s General Electric started selling
  - World's first genetic algorithm product

http://en.wikipedia.org/wiki/Genetic_Algorithm

History Continued V

- 1989 Axelis Inc.
  - World's second GA product
    - First for desktop computers

Methodology

- Definition of Genetic Algorithm
  - Computer simulation
    - Population optimization evolves towards better solution
- What does a Genetic algorithm need?
  - Genetic representation
  - Fitness Function

http://en.wikipedia.org/wiki/Genetic_Algorithm

Methodology Continued

- How it is implemented?
- What is the fitness function?
  - Defined genetic representation
  - Measures quality represented solution
  - Always problem dependent
  - Representation of a solution might be array of bits

http://en.wikipedia.org/wiki/Genetic_Algorithm

Process Initialization/Selection

- How does it get initialized?
  - Individual solutions randomly generated
    - From initial population
- What is the fitness based process?
  - Fitter solutions (as measured by fitness function)
    - Typically more likely to be selected

http://en.wikipedia.org/wiki/Genetic_Algorithm

Process Reproduction

- Mutation (asexual)
  - Low enough probability
- Sexual (crossover)
  - Describe a simple crossover

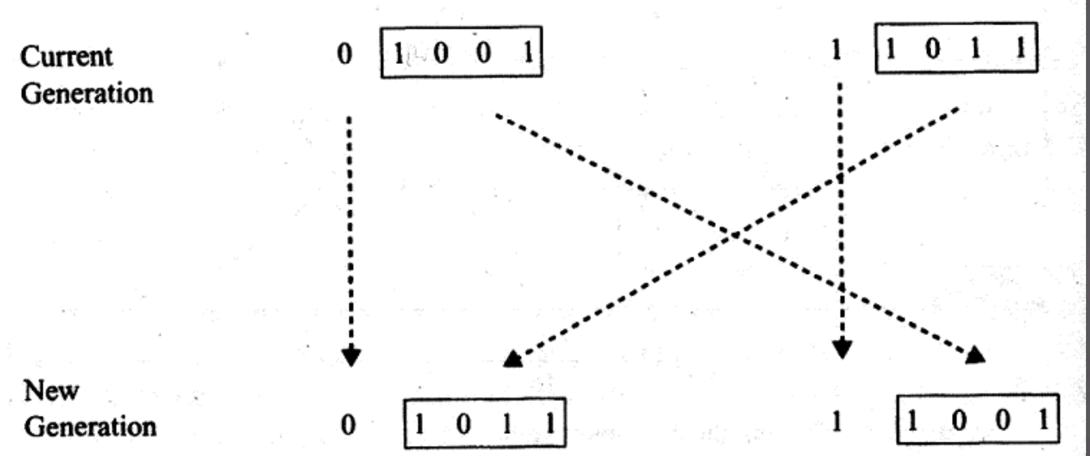

Figure 6.2 Performing crossover at locus two on the first two parents.

Process Termination

- Solution found satisfies minimum criteria
- Fix number of generations reached
- Allocated budget reached
- Fitness solutions reached a plateau
- Manual Inspection
- Combinations of the above

http://en.wikipedia.org/wiki/Genetic_Algorithm

Pros Observations

- Often locate good solutions
- This is an effective heuristic when dealing with a very large solution space
- Mutation introduces new information gene pool
  - Protects against converging too quickly to local optimum

http://en.wikipedia.org/wiki/Genetic_Algorithm

Cons Observations

- Time Delay
- Tend to converge towards local points
  - Rather then global points
- Operate dynamic data sets is difficult
  - May prevent early coverage towards solution

http://en.wikipedia.org/wiki/Genetic_Algorithm

Cons Continued Observations

- Specific optimization problems
  - Simpler optimization algorithms
    - Better solutions than genetic algorithms
- Cannot effectively solve problems which only the fitness measure is right/wrong
  - No way to converge on solution

http://en.wikipedia.org/wiki/Genetic_Algorithm

About Algorithm

- Fitness function is important factor for speed and efficiency of the algorithm
- Selection is important genetic operator
  - Importance crossover versus mutation

http://en.wikipedia.org/wiki/Genetic_Algorithm

Example

- Here is a curve for genetic algorithm

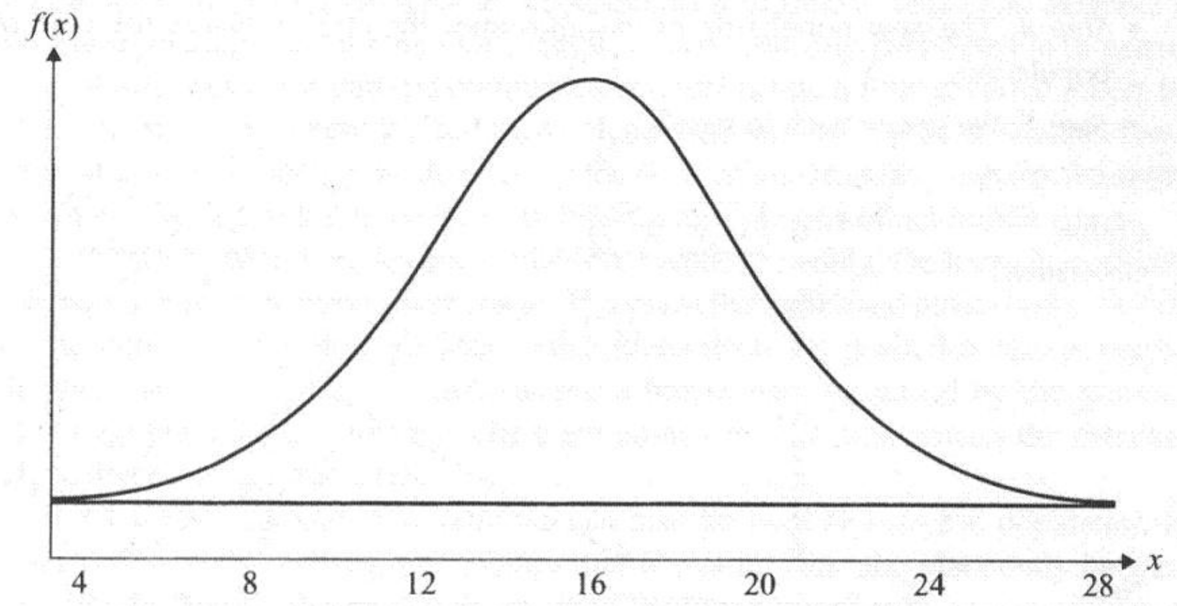

Figure 6.1 Finding the maximum value of the normal **(16, 4)** distribution.

Description of the Curve

SIMPLE EXAMPLE OF A GENETIC ALGORITHM AT WORK

Let's examine a simple example of a genetic algorithm at work. Suppose that our task is to find the maximum value of the normal distribution with mean $\mu = 16$ and standard deviation $\sigma = 4$ (Figure 6.1). That is, we would like to find the maximum value of

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[\frac{-1}{2\sigma^2}(X-\mu)^2\right] = \frac{1}{\sqrt{2\pi}\,(4)} \exp\left[\frac{-1}{2(4)^2}(X-16)^2\right]$$

Steps of GA

We allow X to take on only the values described by the first five binary digits, that is, 00000 through 11111, or 0 to 31 in decimal notation.

First Iteration

- *Step 0: Initialization.* We define the *crossover rate* to be $p_c = 0.75$ and the *mutation rate* to be $p_m = 0.002$.
- *Step 1:* Our population will be a set of four chromosomes chosen randomly from the set 00000 – 11111. So $n = 4$ and $l = 5$. These are *00100* (4), *01001* (9), *11011* (27), and *11111* (31).
- *Step 2:* The fitness $f(x)$ is calculated for each chromosome in the population (Table 6.1).
- *Step 3:* Iterate through the following steps until n offspring have been generated.
  - *Step 3a: Selection.* We have the sum of the fitness values equal to

$$\sum_i f(x_i) = 0.001108 + 0.021569 + 0.002273 + 0.000088 = 0.025038$$

Then the probability that each of our chromosomes will be selected for parenthood is found by dividing their value for $f(x)$ by the sum 0.025038. These

Steps Continued

are also shown in Table 6.1. Clearly, chromosome *01001* gets a very large slice of the roulette wheel! The random selection process gets under way. Suppose that chromosome *01001* and *11011* are selected to be the first pair of parents, since these are the two chromosomes with the highest fitness.

- *Step 3b: Crossover*. The locus is randomly chosen to be the second position. Suppose that the large crossover rate of p_c, 0.75, leads to crossover between *01001* and *11011* occurring at the second position. This is shown in Figure 6.2. Note that the strings are partitioned between the first and second bits. Each child chromosome receives one segment from each of the parents. The two chromosomes thus formed for the new generation are *01011* (11) and *11001* (25).
- *Step 3c: Mutation*. Because of the low mutation rate, suppose that none of the genes for *01011* or *11001* are mutated. We now have two chromosomes in our new population. We need two more, so we cycle back to step 3a.
- *Step 3a: Selection*. Suppose that this time, chromosomes *01001* (9) and *00100* (4) are selected by the roulette wheel method.
- *Step 3b: Crossover*. However, this time suppose that crossover does not take place. Thus, clones of these chromosomes become members of the new generation, *01001* and *00100*. We now have $n = 4$ members in our new population.

- *Step 4*. The new population of chromosomes therefore replaces the current population.
- *Step 5*. We iterate back to step 2.

Data of the First Steps

TABLE 6.1 Fitness and Probability of Selection for Each Chromosome

Chromosome	Decimal Value	Fitness	Selection Probability
00100	4	0.001108	0.04425
01001	9	0.021569	0.86145
11011	27	0.002273	0.09078
11111	31	0.000088	0.00351

Steps of Second and Data

TABLE 6.2 Fitness and Probability of Selection for the Second Generation

Chromosome	Decimal Value	Fitness	Selection Probability
00100	4	0.001108	0.014527
01001	9	0.021569	0.282783
01011	11	0.045662	0.598657
11001	25	0.007935	0.104033

Second Iteration

- *Step 2:* The fitness $f(x)$ is calculated for each chromosome in the population (Table 6.2).
  - *Step 3a: Selection.* The sum of the fitness values for the second generation is $\sum_i f(x_i) = 0.076274$, which means that the average fitness among the chromosomes in the second generation is three times that of the first generation. The selection probabilities are calculated and shown in Table 6.2.

We ask you to continue this example in the exercises.

Pseudo-code algorithm

- 1. Choose initial population
- 2. Evaluate the fitness of each individual
  - In the population
- 3. Repeat
  - 1. Select best-ranking individuals in the population

http://en.wikipedia.org/wiki/Genetic_Algorithm

Pseudo-code algorithm Continued

- 3. Repeat Continued
  - 2. Breed new generation through crossover and mutation (genetic operations)
    - Give birth to offspring
  - 3. Evaluate the individual fitnesses of the offspring
  - 4. Replace worst ranked part of population with offspring
- 4. Until termination

http://en.wikipedia.org/wiki/Genetic_Algorithm

Conclusion

- History Reviewed
- Methodology Reviewed
- Processed Reviewed
- Pros and Cons Discussed
- How it works
- Example and Pseudo-code of the algorithm

References

- Genetic Algorithm, (n.d.), Retrieved March 22, 2008, from http://en.wikipedia.org/wiki/Genetic_Algorithm
- Larose, Daniel T., Data Mining Methods and Models, U.S.A.: John Wiley & Sons, Inc., 2006

Questions