Session #62 TS: Complexity Science: Genetic Algorithm Applications to Actuarial Problems

Brian Charles Grossmiller, FSA, FCA, MAAA
David L. Snell, ASA, ACS, ARA, CLU, ChFC, FLMI, MCP, MAAA

Primary Competency
Technical Skills & Analytical Problem Solving
Why do we call these Genetic Algorithms?

They mimic our current knowledge of genetics.

We have trillions of cells.

DNA represents a blueprint for a cell.

It is used to generate copies.

The actual process involves proteins and lots of other biological terms …
Over time - a long time - we evolve … taller, less hairy, and more adept at making tools.

Very basic genetics

- Humans have about 25,000 genes made from A-T and C-G pairs
- DNA strand – double helix of two strands – each about 1.8 meters
- In meiosis, the strand separates into 46 chromosomes (in 23 pairs)
- Alleles (forms of a gene) help determine physical or behavioral traits
Does the size of a Genome determine species dominance?

NO! – It’s all in how you use it!

- Human, 200,000,000 3.2Gb
- Lungfish, 130,000,000,000 130Gb
- "Amoeba," 670,000,000,000 670Gb

Where is the Research being done on Genomes?

All over the world; but especially …

National Institutes of Health (NIH)

“This is why I go to work!”
RMS – rhabdomyosarcoma
-Frederick S. Huang, MD

The Genome Institute at Washington University
Genetic Algorithms
Meet Amoeba Robby

Robby is only able to do 7 actions:
0-Move North 1 square
1-Move South 1 square
2-Move West 1 square
3-Move East 1 square
4-Stay put
5-Pick up a can
6-Make random move (from 0-5)

Our first Robby is blind!

Story of Robby simplified from *Complexity: A Guided Tour* by Melanie Mitchell, Ph.D.

Genetic Algorithms – 14-JUN-2012 Dave Snell and Brian Grossmiller

Genetic Algorithms (cont.)

Robby actions can be represented by a strategy string, like ‘35265’
(move East, pickup, move South, random, pickup)

In 100 steps, what strategy would you use to pick up as many randomly placed cans as possible?
You get 10 points for each can.
Genetic Algorithms (cont.)

Robby actions can be represented by a strategy string, like ‘35265’
(move East, pickup, move South, random, pickup)

In 100 steps, what strategy would you use to pick up as many randomly placed cans as possible?
You get 10 points for each can.

• Oh yeah! … you lose 5 points every time you hit a wall.

Genetic Algorithms – 14-JUN-2012 Dave Snell and Brian Grossmiller

7^{100} > 10^{84} so trying all possibilities might take too long

… compare to around 10^{17} seconds since Big Bang billions of years ago

Hill Climbing … works really well when there is only one hill

Mount Everest?

Yak, yak, yak

Genetic Algorithms – 14-JUN-2012 Dave Snell and Brian Grossmiller
Genetic Algorithms (cont.)

- Randomly assign the gene string for 100 robots.
- Run them through 100 tests with cans scattered randomly.
- Score each robot and eliminate losers (survival of the fittest).
- Assign mating rights to survivors per test scores.
- Pick parents, and pair them off to produce the next generation (splitting gene string at arbitrary point).

Parent F (52.80 points): 54335351253404 3151421536015206255151153145155663
Parent M (45.60 points): 54335351525534 6332421536042163625515153145155625
Child A: 54335351253404 6332421536042163625515153145155625
Child B: 54335351525534 315142153601520652551511513145155663

- Randomly mutate a few genes (actions)
- Repeat for 100 (or 1,000 or 10,000) generations

Randomly assign the gene string for 100 robots.
Run them through 100 tests with cans scattered randomly.
Score each robot and eliminate losers (survival of the fittest).
Assign mating rights to survivors per test scores.
Pick parents, and pair them off to produce the next generation (splitting gene string at arbitrary point).

Parent F (52.80 points): 54335351253404 3151421536015206255151153145155663
Parent M (45.60 points): 54335351525534 6332421536042163625515153145155625
Child A: 54335351253404 6332421536042163625515153145155625
Child B: 54335351525534 315142153601520652551511513145155663

Randomly mutate a few genes (actions)
Repeat for 100 (or 1,000 or 10,000) generations

Watch Robby pass SOA Life Contingencies exam (MLC)?
Watch evolution in action
Demonstration
... and workshop exercise – Robby

Genetic Algorithms – 14-JUN-2012 Dave Snell and Brian Grossmiller

Genetic Algorithms
(How twins are made)

Genetic Algorithms – 14-JUN-2012 Dave Snell and Brian Grossmiller
**Actuarial Qualification!**

In a real world genetic algorithm, you might not directly assign the gene string actions. You could either create or virtualize a case table covering all conditions and then randomly assign the case actions.

```vbscript
Function fn_CaseNumber(R As Integer, C As Integer) As Integer
    'In an example world where [VisionRadius]=1 and [StatesPerCell]=3, there are three possible conditions for a square (Empty=0, Can=1, Wall=2) and 5 squares visible (Current=0, East=1, South=2, West=3, North=4) thus, there are 3^5 combinations in the set of possibilities.
    'Suppose we want the action for the following combination (current cell, then clockwise):
    '   Current (exponent=0), East (exponent=1), South (exponent=2), West (exponent=3), North (exponent=4)
    '    CAN=1      CAN=1 EMPTY=0 EMPTY=0 WALL=2
    'This would correspond to case number 85 in the virtual case table, calculated as follows:
    '1*3^0 + 1*3^1 + 0*3^2 + 0*3^3 + 2*3^4 = 1+3+0+0+81 = 85
    fn_CaseNumber= Square(R, C) + 3 * Square(R, C + 1) + 9 * Square(R + 1, C) + 27 * Square(R + 1, C - 1) + 81 * Square(R -1, C)
End Function
```

For this particular 'gene string', case number 85 results in action MoveNorth but there is a wall to the North. Note: The actions for any gene string do not have to make sense! Only the better ones survive and propagate.

Impossible case (all walls) … but that’s OK

---

**Mitosis & Meiosis**

Our Elites correspond to Mitosis

The rest of the new generation are analogous to Meiosis
Meiosis in action - Genetics 0.001

Note that each child gene is from two of the four grandparents: one on Mom’s side, one on Dad’s side (excluding mutations).

Parent genes prior to meiosis - each person gets one set from Dad and another set from Mom.

Genetic Algorithms are not limited to two parents (or even four grandparents)

Your ‘genes’ can have vastly different characteristics and components.

A ‘child’ can have genes drawn from several different parents.
That was fun! … but let’s see a Health Insurance application of Genetic Algorithms.

Brian’s application:
How to find a more cost effective set of health providers and still have adequate coverage.

Problem: Picking a Narrow Provider Panel

• An interesting method to reduce healthcare expenditure is to build a product around a high performance network.
• This problem is complicated by the need to maintain adequate access to a comprehensive panel of provider specialties.
Problem: Picking a Narrow Provider Panel

- Production of a narrow yet adequate panel can pose a daunting challenge.
- It is difficult to devise a good starting point and there are a large number of possible solutions.
- Using a binary inclusion variable (In or Out) the number of possible solutions will grow rapidly as the number of providers increase on the scale of $2^N$.

Scale of Provider Network Combinations

- 100 providers have $2^{100} = 1.27 \times 10^{30}$ combinations.
- 500 providers have $2^{500} = 3.27 \times 10^{150}$ combinations.
- 1,000 providers have $2^{1,000} = 1.07 \times 10^{301}$ combinations.
- For comparison the estimated number of atoms in the observable universe is on the scale of $10^{82}$: [http://www.universetoday.com/36302/atoms-in-the-universe/]
Methods to Evaluate Solutions

• Subsets of an existing network will have historical data on cost and quality available by provider.
• This can be used to build a fitness metric to score providers on a relative scale.

Methods to Evaluate Solutions

• Access standards can be generated from a membership target and a specified number of doctors required per member.
• Coupled with a relative score by provider a Genetic Algorithm can be constructed to search for an optimal solution.
Importance of Clean Provider Data

- A narrow panel may be thin in certain specialties, having specialties correct is crucial to ensure solutions are valid.
- Additional details can be useful such as patient capacity, number of hours worked, hospital admitting privileges.
- Grouping providers together appropriately is critical.
- This may require an enhancement or addition to existing data to be usable.

Construction of Fitness Function

- Relative cost scores can be computed based on historical observed claims, though low volume providers may have credibility issues.
- Scores on quality or referral patterns to critical facilities can also be included.
- If more than one score is included they may need to be subjectively weighted into an overall measure.
Construction of Fitness Function

• Weighting together individual scores by doctor or provider group requires several considerations:
  – Weight by claim cost, member utilization, member capacity, or straight provider count?
  – Weighting can be conducted by specialty with the results combined into an overall score.

Construction of Fitness Function

• A non-exhaustive list of possible weighting methods which are typically available are:
  – Allowed Charges
  – Unique Claimants
  – Provider Count
  – Member Capacity
Construction of Fitness Function

- Different methods of weighting can yield slightly different results:

<table>
<thead>
<tr>
<th>Health System</th>
<th>Relative Score</th>
<th>Allowed $</th>
<th>Unique Claimants</th>
<th>Provider Count</th>
<th>Member Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provider #1</td>
<td>1.10</td>
<td>$1,320,000</td>
<td>1,000</td>
<td>10</td>
<td>3,750</td>
</tr>
<tr>
<td>Provider #2</td>
<td>0.80</td>
<td>$504,000</td>
<td>800</td>
<td>6</td>
<td>1,500</td>
</tr>
<tr>
<td>Provider #3</td>
<td>0.85</td>
<td>$1,275,000</td>
<td>1,500</td>
<td>12</td>
<td>4,200</td>
</tr>
</tbody>
</table>

| Weighted Relative Score | 0.96 | 0.94 | 0.95 | 0.96 |

Construction of Fitness Function

- An estimate of cost or utilization by specialty can be used to construct the overall network score:

<table>
<thead>
<tr>
<th>Provider Type</th>
<th>Available</th>
<th>Required</th>
<th>Selected</th>
<th>Relativity</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1.07</td>
<td>50%</td>
</tr>
<tr>
<td>PCP</td>
<td>50</td>
<td>25</td>
<td>25</td>
<td>0.91</td>
<td>15%</td>
</tr>
<tr>
<td>Specialist</td>
<td>75</td>
<td>30</td>
<td>35</td>
<td>0.95</td>
<td>20%</td>
</tr>
<tr>
<td>Other</td>
<td>25</td>
<td>10</td>
<td>15</td>
<td>0.96</td>
<td>15%</td>
</tr>
</tbody>
</table>

Network Relativity 1.01
Implementation of Algorithm

- As the relative scores are deterministic an optimal solution should be attainable.
- Computational resources may be an issue with larger networks or fitness functions requiring more intensive computations.

Implementation of Algorithm

- Determining an optimal solution will depend on the relative score topology and target network size:

Distribution by Relative Score
But does it work?

- The following example in Excel will demonstrate a Genetic Algorithm implementation.

<table>
<thead>
<tr>
<th>Provider Network Example</th>
</tr>
</thead>
</table>

Demonstration

... and workshop exercise – Health Provider Network
But does it work?

• Overall Network Relativity: 1.00
• GA Produced Relativity: 0.72

(a few days later)

• Overall Network Relativity: 1.00
• GA Produced Relativity: 0.72 0.54
Interpretation of Results

• The resulting optimal network will likely require revision to be made practical.
• Some variables such as local provider politics can be highly subjective and will need to be addressed outside of the GA.
• This will provide a useful starting point for building a narrow panel.

Interpretation of Results

• The fitness function provides a convenient format for evaluating any changes that may be required.
• This can also be used to estimate the scale of the pricing differential of the narrow panel.
Basics to take home with you

- Keep problem **manageable**
- **Define objective**
  - Robby - clean world
  - Provider Network - lower health costs and adequate coverage
- **Define universe for the actors**
  - Robby - 10x10 grid littered with soda cans
  - Provider Network - all available providers, their specialties, their relative costs
- **Define actions** – how actors change the universe
  - Blind Robby – deterministic gene string with 7 possible actions for each gene
  - Robby with vision – adapts to environment and chooses actions (from 7)
  - Provider network – deterministic gene string, simple IN or OUT of network
- **Create evaluation/ranking scheme** (points for cans, or relative cost)
- **Make a reproductive model**
  - Blind Robby and Robby with Vision – 2 parents and a crossover point
  - Provider network - N parents and pseudo-random selection gene-by-gene
- **Validate** (save entire gene string for best set and verify it works)

Genetic Algorithms – 14-JUN-2012 Dave Snell and Brian Grossmiller

Wrap-up, and Feedback

Most actuarial presentations are like mathematics, 1/2 the people can’t understand them …

and the other 2/3 don’t care about them

Good tools and Fun? …

… or did we waste your time?

Let’s Talk!

Genetic Algorithms – 14-JUN-2012 Dave Snell and Brian Grossmiller
Genetic Algorithms: Genetic Algorithm Applications to Actuarial Problems
Dave Snell and Brian Grossmiller