Better Marketing Analytics
Using Genetic Algorithms

Doug Newell
Genalytics Inc.
June 2005
Today’s Session

- Direct Marketing Today
- Predictive Modeling Techniques
- Building Models with Genetic Algorithms
Direct Marketing Challenges

- How can I improve my response rates?
- Who are my customers?
- Can I get that targeting done any faster?
- There must be other names out there?
Analytics Can Be the Answer

- Analytics can provide 25%+ better results but
  - Too time consuming and expensive
  - Requires specialized expertise
  - Difficult to justify for all but the largest campaigns
The Traditional Modeling Methods

Linear Regression

Logistic Regression

Neural Networks

Decision Trees
Direct Marketing (US 2004)
Spending: $47 billion
Volume: 94 billion pieces

Reality: Limited Application of Analytics

**Advanced Analytics**

**“Off the Shelf” Models**

**Selection Criteria**

**Mail Everyone**

**“Big Guns”:** Top 50-100 direct marketers

**“Middle Class”:** Moderate mail volume – models come with data

**“Small Fries”:** Small and medium business – best guess

**“Carpet Bombers”:** No targeting
Genetic Algorithms for Predictive Analytics

<table>
<thead>
<tr>
<th>Gene</th>
<th>Interactions (e.g., x+y, x*y, etc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var1</td>
<td>Var2</td>
</tr>
</tbody>
</table>

Chromosome made up of genes and represents model

<table>
<thead>
<tr>
<th>Model 1</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
<th>V8</th>
<th>V9</th>
<th>...</th>
<th>Vn</th>
<th>i1</th>
<th>i2</th>
<th>i3</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 2</td>
<td>V1</td>
<td>V2</td>
<td>V3</td>
<td>V4</td>
<td>V5</td>
<td>V6</td>
<td>V7</td>
<td>V8</td>
<td>V9</td>
<td>...</td>
<td>Vn</td>
<td>i1</td>
<td>i2</td>
<td>i3</td>
<td>...</td>
</tr>
<tr>
<td>Model 3</td>
<td>V1</td>
<td>V2</td>
<td>V3</td>
<td>V4</td>
<td>V5</td>
<td>V6</td>
<td>V7</td>
<td>V8</td>
<td>V9</td>
<td>...</td>
<td>Vn</td>
<td>i1</td>
<td>i2</td>
<td>i3</td>
<td>...</td>
</tr>
<tr>
<td>Model n</td>
<td>V1</td>
<td>V2</td>
<td>V3</td>
<td>V4</td>
<td>V5</td>
<td>V6</td>
<td>V7</td>
<td>V8</td>
<td>V9</td>
<td>...</td>
<td>Vn</td>
<td>i1</td>
<td>i2</td>
<td>i3</td>
<td>...</td>
</tr>
</tbody>
</table>

Multiple chromosomes in a generation

Genalytics' patent-pending process
Begin by Creating a Set of Random Models

<table>
<thead>
<tr>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
<th>V8</th>
<th>V9</th>
<th>...</th>
<th>Vn</th>
<th>i1</th>
<th>i2</th>
<th>i3</th>
<th>...</th>
</tr>
</thead>
</table>

Full Chromosome Represents All Data Variables

Model 1

| V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | V9 | ... | Vn | i1 | i2 | i3 | ... |

Model 2

| V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | V9 | ... | Vn | i1 | i2 | i3 | ... |

Model 3

| V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | V9 | ... | Vn | i1 | i2 | i3 | ... |

Model n

| V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | V9 | ... | Vn | i1 | 12 | i3 | ... |

Model chromosomes use only selected variables
Then Evaluate Fitness for all Models

<table>
<thead>
<tr>
<th>M1</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
<th>V8</th>
<th>V9</th>
<th>...</th>
<th>Vn</th>
<th>i1</th>
<th>i2</th>
<th>i3</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M2</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
<th>V8</th>
<th>V9</th>
<th>...</th>
<th>Vn</th>
<th>i1</th>
<th>i2</th>
<th>i3</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M3</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
<th>V8</th>
<th>V9</th>
<th>...</th>
<th>Vn</th>
<th>i1</th>
<th>i2</th>
<th>i3</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M4</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
<th>V8</th>
<th>V9</th>
<th>...</th>
<th>Vn</th>
<th>i1</th>
<th>i2</th>
<th>i3</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M100</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
<th>V8</th>
<th>V9</th>
<th>...</th>
<th>Vn</th>
<th>i1</th>
<th>i2</th>
<th>i3</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.75</td>
</tr>
</tbody>
</table>

Most fit models are more likely to survive and be selected for breeding
Mating “Most Fit” Models

Mating:

... becomes

11010101010010101010110101010101010101010...
00000101010111101001010010111010101001011...

110101010100101010110101010101010101010...
00000101010111101001010010111010101001011...

Genalytics
Probability of Mating

- Random number “roulette wheel” selects pairs for breeding
- Proportional to fitness
- Higher fitness are more likely to be selected

Genalytics
Multiple Ways to Test Model “Fitness”

- Percent correctly classified
- Pearson’s Linear Correlation
- R-Squared
- Lift
- Upper Lift
- Max K-S
- Error-Squared
Putting it all Together

\[
f_n = 0.277(\text{AGE}) + 5.636(\text{QUANTITY}) - 7.112(\text{ZIP}) + \ldots
\]

Training Data 70%

Score and Rank

“Fitness”

Generation One

Model 1  Model 2  Model 3  Model 4  Model 5  Model 6

Generation Two

Model 7  Model 8  Model 9  Model 10  Model 11  Model 12

Generation (n)

Model 13  Model 14  Model 15  Model 16  Model 17  Model 18

Fitness Accuracy Scale

Model (n)

Low

High
Genetic Algorithm
Advantages
Work with More Data

- **Traditional Approaches**
  - Start with large number of variables
  - Univariate analysis
  - Data reduction
  - Model with “best” 25-50 variables
  - Do you have the most predictive attributes?

- **Genetic Approach**
  - Start with large number of variables
  - No univariate analysis
  - No data reduction
  - Use all variables in software (hundreds or thousands)
  - Attributes with most predictive power thrive
  - Resulting in 10% better predictions
Faster Data Preparation

- Automates many data prep steps
  - Transformations (sq, sqrt, exp, log, etc.)
  - Outlier trimming
  - Missing value substitution
  - Interaction detection
  - Variable selection
  
  Level of human control over the process can be tailored to the situation
More Time for Deeper Exploration

Traditional Regression Based Approach

- Clean & Recode
- Data reduction
- Univariate Analysis
- Model fitting
- Validate
- Implement

Combined GA & Regression

- Load data
- Overnight processing
- Hands-off

- 1-2 hours
- Hands-off

- 2-4 hours
- Validate
- Hand-fit results
- Implement

- 5 to 10 hours

20 to 40+ hours hands on time
Case Study
Leading Financial Services Provider

- Environment
  - Increasing competition
  - Bottleneck in analytics
  - Finite number of analysts
  - Aggressive growth targets

- Business Goals
  - Management wants:
    - Higher response rates
    - Fewer charge-offs
    - Quickly test new ideas
    - Faster model turnaround
Challenge 1: New Market Opportunity

• Situation:
  – Needed further justification to expand new market opportunity

• Hurdle:
  – Quickly build response models against prospect database with over 500 variables
The Results

- Genetic Algorithm:
  - Multiple models in days
  - Evaluated all variables
  - Required one day for hand-fitting model

- Results
  - New business drivers
  - 20% increase in accuracy
  - 20% better response rate
  - Realized $2.5M ROI
The Results are Clear

Genalytics’ models show improvement over clients existing models

*Results based on comparisons with our largest clients*
Genalytics Prospect

- Easy to use, Web-based "wrapper" around analytics software
Contact Information

Genalytics, Inc.
95 Parker St.
Newburyport MA 01950
(978)465-6373
www.genalytics.com
Email: info@genalytics.com