An Introduction to Predictive Analytics
December 2014

Jean-Marc Fix, FSA, MAAA
Vice President, Research and Development
Agenda

- What is it?
- What can it do?
- Uses in life insurance
- Modeling
- Terminology and concepts
- Practical Issues
- Ethical Issues
- Getting started
Goal of Predictive Analytics

- Predict the *customer’s* action based on data about his or her past *behavior*

- Still need to understand the implications to your business and the (economic) environment

- Still need to act!
What is Predictive Analytics?

- A combination of techniques
  - Modeling
  - Data mining
  - (Machine learning)
- Strong statistical component based on regression analysis
- Computer intensive
  - To collect the data
  - To extract the data
  - To validate the data
  - To combine data sources
  - To run models
What can it do?

- Slice bread?
- Nothing
- It predicts
- The *doing* is up to you
What can it predict?

- Everything…
- …based on the past
- Best at predicting a single action based on many inputs
Where is it used?

- Sports (Moneyball by Michael Lewis)
- Credit scores (FICO)
- Netflix movie suggestion system (see also Netflix prize in Wikipedia)
- Auto insurance
Top 10 Google Scholar results for “uses of predictive modeling”

- Predictive modeling of the seismic cycle of the greater San Francisco Bay region
- Predictive modeling of surface roughness and tool wear in hard turning using regression and neural networks
- The use of predictive modeling techniques for optimal exploitation of spatial databases: a case study in landslide hazard mapping with expert system-like methods
- The commercial use of segmentation and predictive modeling techniques for database marketing in the Netherlands
- A framework for predictive modeling of anatomical deformations
- The application of predictive modeling techniques to landslides induced by earthquakes: the case study of the 26 September 1997 Umbria–Marche earthquake (Italy)
- Predictive performance and scalability modeling of a large-scale [hydrodynamic] application
- Predictive statistical models for user modeling
- A comprehensive framework for predictive modeling of negative bias temperature instability
- Predictive modeling of the shelf life of fish under nonisothermal conditions
USES IN LIFE INSURANCE
Uses in Life Insurance

- Who will buy and what will they buy
- Which policyholder will lapse early
- What prospective agent to recruit
- Who will die early/live long
- Which claim is suspect
- How much (economic) reserve should you hold
- Who should you hire
- How much risk is involved in this investment
Automobile vs. Life Insurance

<table>
<thead>
<tr>
<th></th>
<th>Auto</th>
<th>Life</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time horizon</td>
<td>Annual</td>
<td>Decade</td>
</tr>
<tr>
<td>Claims</td>
<td>Multiple</td>
<td>Single</td>
</tr>
<tr>
<td>Frequency of claims</td>
<td>Frequent</td>
<td>Rare</td>
</tr>
<tr>
<td>Severity</td>
<td>Variable, generally low</td>
<td>Fixed, generally high</td>
</tr>
</tbody>
</table>
MODELING
Model Design Step 1: Data Source

- Where is the data?
Input=Data

- Traditional
- Non-traditional (innovative or “out there” depending on your personal bias)
Traditional Data

- Internal
- Vendor data
  - Labs
  - MIB
  - MVR
  - Prescription check
- Industry: SOA, LIMRA,…
- Reinsurance
Data Collection is not a Byproduct

- Purposeful
- High quality
- Understand the limitations
- Address the limitations
- Secure
- Compliant
- Will not happen to get data by accident
- Defined project to increase data quality, connectivity and usability
- Data management area
Non-traditional Data

- Consumer data
  - Producers: Google, Facebook
  - Data brokers: Axiom, Equifax, LexisNexis, Catalina Marketing

- Data aggregates at what level:
  - Not at all aggregated: individual
  - Household
  - Group of 8-25 household
  - Postal carrier route 100-2500 households
  - ZIP code (about 40,000)
Data Tradeoffs

- Accuracy
- Speed
- Cost
- Acceptability
  - Underwriters
  - Management
  - Regulators
  - Agents
  - Prospects
Big Data

- Scary: [http://opendata.zeit.de/widgets/dataretention/](http://opendata.zeit.de/widgets/dataretention/)
- Beyond “1984” ‘s wildest dreams: most individuals carry voluntarily with them localizers that will ping their location throughout the day and will give away their mother’s maiden name in exchange for a chance at a $0.50 coupon
Who Remembers the Kilobyte?

<table>
<thead>
<tr>
<th>Year</th>
<th>Cost per gig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>$437,500</td>
</tr>
<tr>
<td>1985</td>
<td>$105,000</td>
</tr>
<tr>
<td>1990</td>
<td>$11,200</td>
</tr>
<tr>
<td>1995</td>
<td>$1,120</td>
</tr>
<tr>
<td>2000</td>
<td>$11</td>
</tr>
<tr>
<td>2005</td>
<td>$1.24</td>
</tr>
<tr>
<td>2010</td>
<td>$0.09</td>
</tr>
<tr>
<td>2013</td>
<td>$0.05</td>
</tr>
<tr>
<td>2014</td>
<td>$0.03</td>
</tr>
</tbody>
</table>

1981 Apple 5MB for $3,500
1989 Western Digital 20MB for $900
1996 Quantum 2.50GB for $440
2000 Maxtor 41GB for $250
2005 Seagate 200GB for $140

Today: Toshiba 2TB for $69.99 +shipping
Today: Seagate 4TB for $129 +shipping

Big Storage

- Now the terabyte=1,000 gigs is the “new” consumer unit
- Tomorrow: welcome to the petabyte=1,000,000 gigs, “soon” to be available on a “DVD”
- The WWW is estimated to be 4 zettabytes as of June 2013 (1 zettabyte is a million petabytes)
- Perspective: NSA’s Utah center is estimated to store between 3,000 and 12,000 petabytes
Big Analysis

- Voice recognition, OCR, natural language recognition, face recognition
- Combine
- Analyze
- Understand
- Fast
- Amazon Kinesis, Google BigQuery: the birth of real time analytics
### Top Challenges to BDA Programs

#### Percent considered very/extremely challenging

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of financial resources</td>
<td>58%</td>
</tr>
<tr>
<td>Getting executive buy-in</td>
<td>50%</td>
</tr>
<tr>
<td>Accessing data in legacy systems</td>
<td>50%</td>
</tr>
<tr>
<td>Don't know enough about it</td>
<td>37%</td>
</tr>
<tr>
<td>Building business case/value</td>
<td>34%</td>
</tr>
<tr>
<td>Lack of staff/skill sets</td>
<td>34%</td>
</tr>
<tr>
<td>Prioritization</td>
<td>21%</td>
</tr>
<tr>
<td>Confidentiality/privacy</td>
<td>13%</td>
</tr>
<tr>
<td>Other</td>
<td>16%</td>
</tr>
</tbody>
</table>

Source: *The Big Picture: Big Data Analytics in Financial Services*, LIMRA, 2014, used with kind permission
Model Design Step 2: Variable Selection

- Old vs. new data element
- Data level of detail
- Analysis
- Variable selection
Incremental Variable Selection

- Univariate analysis variable vs. output
- Use the most impactful first then add the second most meaningful one. How much better is the model?
- Establish a list of target variables
- Review the targets with all constituencies for acceptability and buy in
Model Design Step 3: Modeling

- Build model
- One possible way:
  - Pick variable having the most impact on the outcome
  - Add second most significant variable
  - Determine weight maximizing the impact
  - Add third…rinse and repeat
Model Design Step 4 & 5

- Validate to another data set
- Calibrate to your need
Issues

- Unexplainable results
- Scarcity of the result variable if looking at death
- Long term nature of the “prediction”, especially of death
- Modeling a proxy variable: modeling the underwriting decision
Model Design Step 6: Revisit

- Review the results quickly and regularly
- Does your output behaves as the model predicted?
  - May need to use proxies, especially for death
- Recalibrate and adjust as needed
TERMINOLOGY AND CONCEPTS
Concepts

- Data mining
  - Rich data
  - Cross dependency/interaction
- Modeling
  - Linear regression
  - Generalized Linear Model
  - Link function
  - Validation
  - Model selection: Akaike Information Criterion
- Variables
  - Lift
  - Fit and overfit:
  - Noise, error distribution
Linear Regression
What’s Next: Overfitting
What is a Good Model?

- Parsimony of model
- Flexibility/adaptability
- Correlation is all that is needed, not causation
- Lead to actionable results or insights
PRACTICAL ISSUES
Practical issues

- Data management
- Statistical know-how
- Recruiting
- Long-term expertise
- Impact on existing departments:
  - Underwriting
  - Marketing/Distribution
  - Pricing (assumption setting)
  - Reserving
  - Financial modeling consistent with predictive modeling
- Resource intensive
- Danger of being a late adopter
ETHICAL ISSUES
Restrictions

- Fair Credit Reporting Act (US)
  - Consumer
    - must be told if info in file used against him
    - know what is in the file
    - can dispute incomplete/inaccurate info
  - Reporter
    - must delete/correct info

- Other legal (unfair discrimination)
Ethical Issues

- Unwitting inclusion of proxies for prohibited variables
- Encouraging lapses of unprofitable customers: generally not in customer’s best interest
- Redefining underwriting
- Managing reputational risk: transparency and consistency of results
GETTING STARTED
Resources and Next Steps

- **Read-up**
  - SOA soon to be released research
  - LIMRA/LOMA research
  - On the Risk article by Mark Dion

- **Attend seminar**
- **Practice on a small scale using data you own and understand: get a pilot started**
- **Is there in-house expertise?**
- **Consultants**
THE SPEAKER
Jean-Marc is Vice President, Research and Development at Optimum Re. He is responsible for the evaluation of new concepts involving product development or underwriting. His special areas of interest are all things mortality and critical illness both in the US and Canada.

Prior to joining Optimum in 1997, Jean-Marc has worked at a number of reinsurer and direct life insurance companies, mostly in the area of product development.

Jean-Marc is a frequent speaker at life insurance, underwriting and medical directors meetings and is currently involved with a number of industry activities, including several predictive modeling research oversight groups and:

- Joint AAA/ SOA team on guaranteed and simplified issue mortality
- Joint AAA /SOA Underwriting Criteria Team
- SOA Living to Age 100 Symposium Program Committee
- SOA Reinsurance Section Research Team
- SOA Longevity Advisory Group

Jean-Marc received a BA in Mathematics from Whittier College in California and can be contacted at (214) 528-2020 or jean-marc.fix@optimumre.com