Predictive Modeling and Dental Fraud Detection

November 22, 2007

Jonathan Polon FSA
Barry Senensky FSA FCIA MAAA
Claim Analytics Inc.
www.claimanalytics.com
Agenda

- Introduction to Predictive Modeling
- Overview of Dental Fraud Paper
- Predictive Modeling vs Classical Stats
- Supervised vs Unsupervised Learning
- Predictive Modeling for Fraud Detection
2007 General Meeting

Assemblée générale 2007
Introduction to Predictive Modeling
## Computer Performance

<table>
<thead>
<tr>
<th>Measure</th>
<th>IBM 7094 c. 1967</th>
<th>Laptop c. 2004</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor Speed (MIPS)</td>
<td>0.25</td>
<td>2,000</td>
<td>8,000-fold increase</td>
</tr>
<tr>
<td>Main Memory</td>
<td>144 KB</td>
<td>256,000 KB</td>
<td>1,778-fold increase</td>
</tr>
<tr>
<td>Approx. Cost ($2003)</td>
<td>$11,000,000</td>
<td>$2,000</td>
<td>5,500-fold decrease</td>
</tr>
</tbody>
</table>
What is a Predictive Model

• A Predictive Model is a model which is created or chosen to try to best predict the probability of an outcome

• Two “ERA’s” of Predictive Models
  – Traditional
  – Modern
Traditional Predictive Models

• Generally type used by Actuaries
• Examples include
  – Mortality Studies
  – Lapse Studies
  – Morbidity Studies
Characteristics of Traditional Predictive Models

- Relatively low usage of computer power
- Usually require many assumptions
- May need large amounts of data to maintain credibility
- Relatively easy to understand
Modern Predictive Models

• Have been around 40+ years
• Used extensively in industry
• Applications include
  – Credit Scores
  – Credit Card Fraud Detection
  – Stock Selection
  – Mail sorting
  – Hot dogs and Hamburgers
Examples of Modern Predictive Modeling Methods

- Classification and Regression Trees
- Neural Networks
- Genetic Algorithms
- Stochastic Gradient Boosted Trees
- Clustering Techniques
  - K Means
  - Expectation Maximization
Characteristics of Modern Predictive Models

- Intense usage of computer power
- Require few assumptions
- Can maintain credibility with lesser of amounts of data
- Models can be black boxes
Why aren’t Actuaries building modern predictive models?

• Life Insurance Industry is conservative and slow to change
• Not a traditional actuarial tool
• The times are changing!
  – Especially P&C Actuaries
• It’s only a matter of time!
  – It just makes too much sense!
  – Innumerable applications to help solve insurance problems
Applications of Modern Predictive Models to Insurance Problems

- Traditional
  - Better understanding of experience
  - Better pricing
  - Better reserving
  - Better underwriting

- Non-traditional
  - More objective claims management
  - Improved fraud detection
Overview of Dental Fraud Research Paper
Dental Fraud Research

- Started with approximately 200,000 dental claims records
- Used modern predictive modeling techniques to identify dentists with “atypical” insurance claims practices
- Project was very successful
Dental Fraud Research (Cont’d)

- Research Paper titled “Dental Insurance Claims Identification of Atypical Claims Activity”
- Published April 2007 and available on CIA Website and our website www.claimanalytics.com
Predictive Modeling Vs Classical Statistics
## What is Predictive Modeling?

<table>
<thead>
<tr>
<th>Action</th>
<th>Classical Statistics</th>
<th>Predictive Modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyse historic data</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Identify &amp; quantify relationships between inputs and outcomes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Apply this learning to predict outcomes of new cases</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

So what’s the difference?
The essentials haven’t changed.

But many things are different.
# Exploiting Modern Computing

<table>
<thead>
<tr>
<th>Electronic computing</th>
<th>Classical Statistics</th>
<th>Predictive Modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Approach</strong></td>
<td>Predates it</td>
<td>Exploits it</td>
</tr>
<tr>
<td></td>
<td>Computationally efficient</td>
<td>Computationally intense</td>
</tr>
<tr>
<td></td>
<td>eg, tables</td>
<td>Test complex relationships</td>
</tr>
<tr>
<td></td>
<td>eg, common distributions</td>
<td>Use a numeric approach</td>
</tr>
</tbody>
</table>
Exploiting Computing Power

To Build a Predictive Model:

- Analyst specifies desired form of the model
- Applies iterative, numerical approach to optimize weights
- Common approaches include: gradient descent, competitive learning and exhaustive search
Predictive Modeling Summary

Primary Advantage

• Can quantify greater complexity of relationships between input and outcome
• Can result in much better accuracy than traditional techniques

Primary Concern

• Models can be difficult to interpret
Common Modeling Techniques

• Generalized Linear Models
• Neural Networks
• Genetic Algorithms
• Stochastic Gradient Boosted Trees
• Random Forests
• Support Vector Machines
Supervised Vs Unsupervised Learning
Supervised Learning

- **Known** outcome associated with each record in training dataset
- **Objective**: build a model to accurately estimate outcomes for each record
- **Eg**, Predicting claim incidence rates or severities based on past experience
- Commonly referred to as **predictive modeling**
Unsupervised Learning

- **No known** outcome associated with any record in training dataset
- **Objective:** self-organization or clustering; finding **structure** in the data
- **Eg,** Grocery stores: define types of shoppers and their preferences
- Insurance fraud is typically unsupervised learning. Because it is **not known** which historic claims were fraudulent.
Dental Fraud Detection
Using
Unsupervised Learning
Dental Fraud Project Overview

- Explore use of pattern detection tools in detecting dental claim anomalies
- 2004 data
- 1,600 Ontario GPs (non-specialists)
- 200,000 claims
Traditional Tools

✓ Rule-based

✓ Strong at identifying claims that match known types of fraudulent activity

✓ Limited to identifying what is known

✓ Typically analyze at the level of a single claim, in isolation
Pattern Detection Tools

• Analyze millions of claims to reveal patterns
  – Technology learns what is normal and what is atypical – **not limited by what we already know**

• Categorize each claim. Reveal dentists with high percentage of atypical claims
  – Immediately highlight **new** questionable behaviors
  – Find new large-dollar schemes
  – Also identify frequent repetition of small-dollar abuses
Methodology – 3 Perspectives

1. **By Claim**  e.g. Joe Green’s semi-annual check-up

2. **By Tooth**  e.g. all work done in 2004 on Joe’s bicuspoid by Dr. Brown

3. **By General Work**  e.g. all general work (exams, fluoride, radiographs, scaling and polishing) done on Joe in 2004 by Dr. Brown
Methodology – 2 Techniques

1. Principal components analysis

2. Clustering
## Methodology – Summary

<table>
<thead>
<tr>
<th></th>
<th>PCA</th>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claim by claim</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tooth by tooth</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>General work</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Principal Components Analysis

- Visualization technique
- Allows reduction of multi-dimensional data to lower dimensions while maximizing amount of information preserved
- Powerful approach for identifying outliers
PCA: Simple Example

Lots of variance between points around both axes
PCA: Simple Example

Create new axes, X’ and Y’: rotate original axes 45°
PCA: Simple Example

In the new axes, very little variance around $Y'$ - $Y'$ contains little "information"
PCA: Simple Example

Can set all Y' values to 0 – ie, ignore Y’ axis
Result: reduce to 1 dimension with little loss of info
PCA: Identifying Atypical Dentists

PCA: Graph of Claims (Visits to the Dentist)
PCA: Identifying Atypical Dentists

1. Begin at the individual transaction level

2. Determine the average transaction for each dentist

3. Graph quickly isolates dentists that are outliers
PCA: Identifying Atypical Dentists

Dentist Average

90th percentile
PCA: Summary

1. Visualization allows for easy and intuitive identification of dentists that are atypical

2. Manual investigation required to understand why a dentist or claim is atypical
Clustering Techniques

• Categorization tools

• Organize claims into several groups of similar claims

• Applicable for profiling dentists by looking at the percentage of claims in each cluster

• We apply two different clustering techniques
  – K-Means
  – Expectation-Maximization
Clustering Example

Clustering Example

Handicap

# Weekly Golf Games

0 2 4 6 8
Clustering Example

![Clustering Example Diagram]

- Handicap
- # Weekly Golf Games

2007 General Meeting
Assemblée générale 2007
Eg: Clusters found – by Claim

1. Major dental problems
2. Moderate dental problems
3. Age > 28, minor work performed
4. Work includes unbundled procedures
5. Minor work and expensive technologies
6. Age < 28, minor work performed
Clusters: Identifying Atypical Dentists

1. Begin at the individual claim level

2. Calculate the proportion of claims in each cluster – in total, and for each dentist

3. Isolate dentists with large deviations from average
Clusters: Identifying Atypical Dentists

Proportion of General Work by Cluster
K-Means Clustering

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Average</th>
<th>Dent A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19%</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>6%</td>
<td>0%</td>
</tr>
<tr>
<td>4</td>
<td>20%</td>
<td>0%</td>
</tr>
<tr>
<td>5</td>
<td>6%</td>
<td>0%</td>
</tr>
<tr>
<td>6</td>
<td>9%</td>
<td>49%</td>
</tr>
<tr>
<td>7</td>
<td>21%</td>
<td>0%</td>
</tr>
<tr>
<td>8</td>
<td>17%</td>
<td>51%</td>
</tr>
</tbody>
</table>
Clustering Techniques: Summary

1. Clustering provides an easy to understand method to profile dentists

2. Unlike PCA, clustering tells why a given dentist is considered atypical

3. Effectively identifies atypical activity at the dentist level
Pilot Project Results

- PCA identified 68 dentists that are atypical
- Clusters identified 182 dentists that are atypical
- 36 dentists are identified as atypical by both PCA and Clusters
- In total, 214 of 1,644 dentists are identified as atypical (13%)
- Billings by atypical dentists were $2.5 MM out of $16.1 MM billed by all dentists (15%)
Examples

Of

Atypical Dentists
Analysis: Dentist A is far beyond norms in clusters 6 and 8; both indicate high charges in ‘general dentistry’

What we discovered: Each of Dentist A’s patients is being billed for at least 30 minutes of polishing and 45 minutes of scaling
Analysis: Dentist B is far beyond norm in cluster 7

What we discovered: Frequent extractions and anesthesia – Dentist B looks like an oral surgeon, yet is a general practitioner
**Analysis**: Very high proportion in Cluster 1, suggesting many high-ticket visits

**What we discovered**: First, Dentist C performs an inordinate amount of scaling. Second, Dentist C has emergency examinations with atypically high frequency.
Analysis: Dentist M has a disproportionate amount of work beyond the 90th percentile of work by all dentists on all teeth.

What we discovered: Dentist M appears to utilize lab work very heavily.
More Atypical Dentists

- Frequent use of panoramic x-rays
- Frequent use of nitrous oxide – including with procedures rarely associated with anesthesia
- Large number of extractions, often using multiple types of sedation
- Very high proportion of claims for crowns and endodontic work
- Individual instruction on oral hygiene provided with very atypical frequency
Dental Fraud Summary

Pattern detection:

• Highly effective in identifying dentists with claim portfolios significantly different from the norm

• Enables experts to quickly identify and focus on those dentists with atypical claims activity
Questions?