The Good, the Bad and the Ugly - Techniques for Portfolio Segmentation

Presented by Nelson Henwood

Finity Commercial Lines Pricing Seminar
4th May 2011
I want you to present to the Board next week outlining your strategy to compete effectively whilst maintaining profitability on these contestable platforms for Package business which we have signed up to.
Today’s Presentation

- The Challenge of Portfolio Segmentation
- A Machine Learning Based Approach
- Portfolio Segmentation Examples
- Two Bits on Data
- Some Ammo for our Friend
- Conclusions
The Challenge of Portfolio Segmentation

- Dissect the book into blocks of business which exhibit similar characteristics
- To enable the implementation of targeted treatment strategies

- Some possible segmentation targets:
  - Profitability
  - Sales performance / growth
  - Price sensitivity
  - Combinations of the above
  - Changes in the above
"[Traditional approaches] require a sophisticated (domain knowledge) user who does not sleep or age. The user must repeatedly formulate (guess) informative queries"
What is Machine Learning?

“...computer algorithms that improve automatically through experience”


- Learner (algorithm) processes data representing past experiences and tries to either:
  - Develop an appropriate response to future data – Supervised Learning
  - Understand the relationships between the data components – Unsupervised Learning

[Lecture Notes, CIS 526, Temple University]
### How is Machine Learning relevant?

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficient</td>
<td>Rapid search through large data files</td>
</tr>
<tr>
<td>Powerful</td>
<td>Sensitive (but not overly so) to performance differentials in target</td>
</tr>
<tr>
<td>Flexibility</td>
<td>Directly address questions at cost or policy level (including multiple sections)</td>
</tr>
<tr>
<td>Interpretability</td>
<td>Segment descriptions in plain English</td>
</tr>
</tbody>
</table>
What does it do?

- Identifies customer segments by uncovering key multivariate, non-linear interactions of variables

- Customer segments are:
  - Mutually exclusive and exhaustive
  - Credible
  - Actionable and easy to communicate

- Validate findings by demonstrating segmentation results on data held back from analysis
How do we do it?

- We use a proprietary software solution called Talon. Finity is the only company holding a license for this software in Australia.

- Talon employs a set of unique supervised learning algorithms based on a combination of statistical and computational techniques. It has been specifically built for insurance applications.

The unique proposition from the use of Talon is its collection of domain specific unique learning methods which have been adapted through many years of research to detect significant, persistent signals in insurance data which cannot be found efficiently with other techniques.
## Analysis Types

| Complete Segmentation |  ● Mutually exclusive and exhaustive segments  
|                       |  ● Plain English description  
|                       |  ● Similar and consistent outcome over target  
|                       |  ● Target examples - loss ratio, profit per policy, frequency, retention etc  
| Best / Worst Segmentation |  ● Focus on extremes of performance across target  
|                          |  ● Describe best / worst x% of book with respect to (loss ratio, say)  
|                          |  ● Plain English description  
| Pure Trend Segmentation |  ● Determines growing / declining portfolio segments ie mix of business changes  
|                         |  ● Written or exposure based  
|                         |  ● Mutually exclusive and exhaustive segments  
|                         |  ● Plain English description |
## Analysis Types (ctd)

<table>
<thead>
<tr>
<th>Analysis Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instability Segmentation</td>
<td>- Identify segments experiencing a change in target outcome&lt;br&gt;- Mutually exclusive and exhaustive segments&lt;br&gt;- Plain English description&lt;br&gt;- Target examples - claim frequency, average size, loss ratio, quote conversion, retention etc</td>
</tr>
<tr>
<td>Joint Objective Segmentation</td>
<td>- 2-way portfolio segmentation focussing on (say) profitability and growth&lt;br&gt;- Eg - unprofitable and growing segments, profitable and declining&lt;br&gt;- Mutually exclusive and exhaustive segments&lt;br&gt;- Plain English description</td>
</tr>
<tr>
<td>Scoring</td>
<td>- Determines a &quot;score&quot; for each policy (1-1000) with high correlation to target&lt;br&gt;- Achieves greater separation than Complete Segmentation&lt;br&gt;- Trade-off is transparency - score vs plain English description</td>
</tr>
</tbody>
</table>
Examples
Some Examples

- Profitability segmentation
  - Small Fleet
  - Commercial Property – ISA Property data for Package

- Monitoring for mix of business changes (Pure trend)
  - Comprehensive Motor

- Combined Profitability / Growth Segmentation
  - Small Fleet
Small Fleet – Complete Segmentation
Background

- Exposure rated fleet portfolio
- Fleets up to 20 vehicles
- Written Australia wide
- Mix of light and medium commercial
- Rating flexibility about average
- Leading market player
- Modelling undertaken at risk level and fleet level
Risk Level – Model Fit

- 13 segments identified
- Large range of relative profitability
  - Lift = 2.1
- Good correlation between training (70%) and validation data (30%) at 97%
Risk - Segment Descriptions

- Segments are 3-4 way combinations of variables
- Only 4 variables used in the segmentation – vehicle type and location are the most powerful discriminators

<table>
<thead>
<tr>
<th>Segment</th>
<th>% of Exposure</th>
<th>Loss Ratio</th>
<th>Vehicle Type</th>
<th>Location</th>
<th>Duration</th>
<th>Excess</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>11% 68%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st #6</td>
<td>9% 83%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad</td>
<td>6% 111%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd #9</td>
<td>12% 112%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ugly</td>
<td>8% 134%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd #8</td>
<td>6% 141%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st #5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<....7 segments in between....>
Fleet Level – Model Fit

- 9 segments identified at fleet level
- Lift is lower than for risk level model (2.1 vs 1.7) – some diversification at fleet level
- Strong correlation from training (70%) to validation (30%) at 94.5%
Fleet – Segment Descriptions

- Fleets are segmented based on combinations of 1-4 variables
- The primary factor driving the segmentation is the fleet composition rating derived from the risk model
- Other factors such as fleet size, duration and broker group also feature

<table>
<thead>
<tr>
<th>Segment</th>
<th>% of Exposure</th>
<th>Loss Ratio</th>
<th>Relativity</th>
<th>Fleet Comp Rating</th>
<th>Vehicle Count</th>
<th>Duration</th>
<th>Broker Group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Good</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st #5</td>
<td>10%</td>
<td>76%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd #2</td>
<td>10%</td>
<td>86%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd #1</td>
<td>11%</td>
<td>87%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bad</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd #9</td>
<td>9%</td>
<td>110%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st #9</td>
<td>9%</td>
<td>115%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ugly</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st #13</td>
<td>13%</td>
<td>129%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Industry Data – Package Property – Complete Segmentation
Background

- ISA Commercial Property database
- Restricted to package business
- Analysis of non-weather claims only
- Merged socio-demographic factors at postcode level
- NSW metro analysed
- Minimum exposure threshold for segmentation set to 20,000 years

- Preliminary results
- Socio demographic factors dominate the segmentation
- More work required on occupation clustering
10 segments identified, each containing at least 8% of the portfolio exposure

Large range of performance differentials (lift ~ 2.5)

Good correlation between training (70%) and validation (30%) at around 93% - some volatility noted
Geographic Segmentation – Sydney Metro
Comprehensive Motor – Trend Segmentation
Background

- Comprehensive Motor
- Call centre and internet distributed
- Medium size
- Mature portfolio growing slowly
- Analysed changes in trend for new business written by segment 3 months either side of a rate review

Key questions:
- What was the change in mix of business written?
- How did that fit with our expectation?
Validation by Random Sampling

- 7 segments were identified by the analysis
- We used 20 repeated random samples of 50% of the data used to validate model results
  - Average correlation ~ 98%
Segment Descriptions

- Segments consist of 2-4 way combinations of variables
- Policy and vehicle factors incorporated

<table>
<thead>
<tr>
<th>Segment</th>
<th>Relative Growth</th>
<th>Vehicle Factor #1</th>
<th>Driver Age</th>
<th>Policy Factor #2</th>
<th>Payment Freq</th>
<th>Vehicle Factor #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative Decline</td>
<td>#2  -12%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#5  -9%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#4  -5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative Growth</td>
<td>#6   7%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#3  30%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<....2 segments in between....>
Business Written and Price Change

New Business Written by Month

Price Change vs Growth

Growth in New Business

Segment 1 — Segment 2 — Segment 3 — Segment 4
Segment 5 — Segment 6 — Segment 7
Fleet Motor – Combined Objective
Partition portfolio on the basis of a joint objective:
- Loss ratio and
- Change in exposure
- New and renewal business
- Trend over 3 accident years examined
Loss ratio validates reasonably well – correlation 92.5%

Correlation of growth validation very high (~98%)
Training vs Validation Results
### Segment Descriptions

- Segment descriptions are relatively simple – 1 to 4 factors involved
- Huge range of loss ratio relativities
- Some clear adverse trends apparent

<table>
<thead>
<tr>
<th>Segment</th>
<th>% of Exposure Pre</th>
<th>% of Exposure Post</th>
<th>Loss Ratio Relativity</th>
<th>Vehicle Age</th>
<th>Policy Factor #1</th>
<th>Policy Factor #2</th>
<th>Geo Factor #1</th>
<th>Geo Factor #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profitable and Declining #2</td>
<td>14%</td>
<td>13%</td>
<td>36%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profitable and Growing #6</td>
<td>11%</td>
<td>13%</td>
<td>77%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unprofitable and Declining #7</td>
<td>10%</td>
<td>9%</td>
<td>141%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unprofitable and Growing #3</td>
<td>12%</td>
<td>14%</td>
<td>129%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13%</td>
<td>17%</td>
<td>145%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2 Bits on Data
Data Preparation – A Key Step

• Data preparation is very important to achieving good results – the machine won’t do this for you!!

• Need to deal with:
  ▶ Weather perils / catastrophes
  ▶ Large claims
  ▶ Variables with a large number of levels eg postcode / occupation
  ▶ Changes in rating / discounting

• As well as the usual noisy, messy stuff such as changes in:
  ▶ The way data is recorded
  ▶ Underwriting standards / excess levels etc
External Data for Non-Weather Losses

- Some important pieces of data for pricing / segmentation:
  - Socio-demographic
  - Positioning data
  - Business information
  - Credit data

- Dealing with occupation:
  - Developing attributes
  - An insurance risk based hierarchy
Some Ammo for our Friend
Package Initiatives – Sharpen the Pricing

- Pricing improvements
  - Improved alignment of pricing to risk
  - Decreased cross-subsidies
- Limitations in the rating structure means that all not issues can be addressed
  - Prioritised changes for development
Package Initiatives - Monitoring

- Analysis to detect changes in mix of business
  - Separately for new business and renewals
  - Assess response of business written to any price changes
- Analysis to detect changes in retention and quote conversion rates
  - Signal regarding movements in price competitiveness
Package Initiatives - Renewal Underwriting

- Specific renewal segmentation which incorporates information learned about a policy once it is on the books eg claiming behaviour, policy changes
- Used to prioritised re-underwriting at renewal

![Graph showing relative COE for different segments](image)

- Target 10% of this segment for non-renewal
- Discretionary renewal discounts
- Automated loadings applied
Targeting New Business

- Articulating target segments to brokers
  - Gathering feedback on competitive positioning
- Monitoring quote conversion to ensure we are winning our share

- All trades excluding plumbers and welders
- North of Harbour

- Take away shops
- West of Homebush

- All North shore locations in business for more than 2 years
Conclusions

- The introduction of contestable platforms increases the need for segment-based portfolio strategies.
- Price will become a more important factor in determining the business that is won.
- Challenging to understand competitive positioning for commercial but can monitor:
  - Changes in mix of business
  - Changes in retention / quote conversion
- It is likely that more frequent revisions to pricing will be necessary.
- The key to success is understanding what is happening at segment level and taking multiple, related actions to address.
This presentation has been prepared for the Finity Commercial Lines Pricing Seminar held on 4 May 2011. Finity Consulting Pty Limited (ABN 89 111 470 270) wishes it to be understood that opinions put forward herein are not necessarily those of Finity and Finity is not responsible for those opinions. The information presented at the seminar was of a general nature and a reader of this presentation must seek their own independent advice before using it for any purpose.