Using Machine Learning Techniques to Enhance Predictive Models

Nelson Henwood, Wei Lin and John Yick

© Finity Consulting, 2010
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]
Machine Learning Algorithms

• Include:
  – Decision Trees
  – Neural Networks
  – Support Vector Machines (SVMs)
  – Gaussian process regression
  – Ridge Regression
  – Nearest Neighbours
  – Partial least squares

• Presentation results are based on Talon algorithms:
  – Supervised learning
  – Proprietary algorithms
  – Fusion of techniques
  – Tailored for insurance data
Ensemble Techniques

• Systems for combining predictions from multiple models to obtain a more powerful predictor than could be obtained from any of the constituent models

• Requires:
  – Prediction diversity
  – Combination rules

<table>
<thead>
<tr>
<th>Ensemble System</th>
<th>How it achieves diversity?</th>
<th>Combination method - Classification / Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bagging</td>
<td>Bootstrap resampling</td>
<td>Majority vote / mean</td>
</tr>
<tr>
<td>Random Forests</td>
<td>Bagging + random selection of predictors</td>
<td>Majority vote / mean</td>
</tr>
<tr>
<td>Boosting</td>
<td>Biased resampling</td>
<td>Weighted average majority vote / weighted mean</td>
</tr>
</tbody>
</table>

How is ML different to current predictive modelling practice? (GLMs)

<table>
<thead>
<tr>
<th>Current Practice</th>
<th>Machine Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔ Assumes that variables are independent unless specifically defined otherwise</td>
<td>✔ Allows data to interact naturally to find the patterns between characteristics within the data</td>
</tr>
<tr>
<td>✔ “Optimal” predictors are based on assumptions</td>
<td>✔ “Data Speaks For Itself”</td>
</tr>
<tr>
<td>✔ Can’t solve what you don’t know</td>
<td>✔ Does not require the user to specify the predictors and interactions to be included in the model - it discovers them</td>
</tr>
<tr>
<td>✔ Large potential volume of interactions results in only a small subset being investigated</td>
<td>✔ Extremely Fast and Efficient</td>
</tr>
<tr>
<td>✔ Pricing cost models are done at a peril level versus a policy or customer level</td>
<td>✔ Can be performed at peril, policy or customer level</td>
</tr>
<tr>
<td>✔ Additional model assumptions required eg linearity, error distribution</td>
<td>✔ No assumptions regarding linearity, assumed error distribution much less important</td>
</tr>
</tbody>
</table>
Signal Validation

Validation by Random Partition
- Training using randomly selected 70% of the data
- Validate on remaining 30%

A More Cautious Strategy
- 3-way split of the data
- Training using say 50%-70% of the data
- Validate using 20% (or random cross-section)
- Use 30% unseen data for post-modelling test
Processes Used in Talon for Analysis

- **Complete Segmentation** – partitions the portfolio into mutually exclusive and exhaustive segments where each segment contains policies that have common attributes and have similar and consistent loss ratio, frequency, severity, loss cost, profit per policy, retention or conversion.

- **Multi-split** – undertakes multiple segmentation runs and incorporates a random element in the splitting process – doesn’t always take the “optimum path”. Can result in different areas of the data being explored and can lower the risk of finding a local minima.

- **Scoring** – a practically continuous segmentation of a portfolio achieved through application of ensemble techniques. Produces greater differentiation than Complete Segmentation with respect to the factor of interest.
Our Business Problem

- Build an accurate predictive model of claims cost for a comprehensive motor portfolio
- Frequency and size models constructed at peril (group) level
- GLMs and Decision Tree techniques (CART) were employed

- Some of our constraints include:
  - Model assumptions:
    - inherent in the GLM – linearity, error distribution
    - typically assume independence of perils and frequency / size
  - Sleeping and ageing

- Can machine learning techniques be used to find consistent signals in the model residuals and thus improve the predictions?
  - Results shown for at-fault collision models and total cost
At-fault Collision models fit “well”

- For frequency we have:
  - Reasonable gains chart
  - Model Pearson residual 38% lower than null model
  - CART won’t split the residual further than shown with 5,000 exposure year threshold
Key Parameters for Segmentation

• Our segmentation target is the actual vs modelled claim frequency / size

• Minimum volume threshold for a segment is a key parameter for segmentation to ensure a credible segmentation
  – Frequency and Cost – exposure
  – Size – number of claims

• One test we undertook was to explore a number of different volume thresholds using a Complete Segmentation

• We examined the degree of segmentation vs persistency of the signal
  – Lift = ratio of prediction (A vs E) in highest to lowest segment
  – Correlation is measured on A vs E across segments between the training (70%) and validation (30%) data
Frequency – Volume Threshold Testing
Size – Volume Threshold Testing

Lift vs Correlation

Change in Gini vs Correlation

Complete Segmentation for Volume Threshold (000)
Selected Volume Threshold
Multi Split Runs

- **Frequency**
  - Correlation vs. Lift
  - Graph shows distribution of data points for individual multi-split runs.

- **Average Size**
  - Correlation vs. Lift
  - Graph shows distribution of data points for complete and selected segmentation.

Legend:
- Gray: Individual Multi-Split Run
- Blue: Complete Segmentation
- Red: Selected Segmentation
Complete Segmentation - Frequency

- A number of segments are identified which show significant and persistent deviation from the model fit.
Frequency Signal Validation

- Signal persists across the validation data – better persistence at the extremes
Frequency Example Segments

- The segments identified are quite complex
- Minimum of 4 variable conditions required to describe
- A vs E shown over training + validation data

<table>
<thead>
<tr>
<th>Segment #</th>
<th>% Exposure</th>
<th>Actual / Model</th>
<th>Segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>8%</td>
<td>0.82</td>
<td>6 way interaction involving vehicle characteristics, NCB level, policy status and others</td>
</tr>
<tr>
<td>4</td>
<td>5%</td>
<td>0.87</td>
<td>7 way interaction involving vehicle characteristics, NCB level, policy status and others</td>
</tr>
<tr>
<td>5</td>
<td>5%</td>
<td>1.07</td>
<td>4 way interaction involving vehicle characteristics, NCB level and policy status</td>
</tr>
<tr>
<td>10</td>
<td>6%</td>
<td>1.09</td>
<td>4 way interaction, including vehicle characteristics, policy duration and NCB level</td>
</tr>
<tr>
<td>15</td>
<td>6%</td>
<td>1.11</td>
<td>4 way interaction, including vehicle characteristics, policy duration and NCB level</td>
</tr>
<tr>
<td>8</td>
<td>7%</td>
<td>1.11</td>
<td>4 way interaction including age, NCB level and others</td>
</tr>
<tr>
<td>16</td>
<td>6%</td>
<td>1.20</td>
<td>4 way interaction including vehicle characteristics, policy duration and others</td>
</tr>
</tbody>
</table>
**Frequency Scoring**

- The scoring run shows a strong persistent signal
- The signal validates quite well across the entire spectrum of mis-fit identified

![Frequency Scoring Diagram](image-url)
Frequency – Scoring vs Segmentation

• Scoring delivers more lift than segmentation – here shown across the validation data
Modification of GLM Prediction - Frequency

- Data divided into deciles based on GLM predicted frequency
- Chart illustrates how the Complete Segmentation and Scoring “pull-apart” the GLM prediction
- Nb: log scale used
Complete Segmentation - Average Size

- Analysis of the average size model also reveals a number of segments with significant deviation from the model fit.
Average Size – Signal validation

- The lift is not as great as for the frequency model
- Persistence of the signal across the validation data is good at the extremes
Modification of GLM Prediction - Size

- The segmentation and scoring also significantly “pull-apart” the GLM prediction across the range of average size.
Cost = modelled frequency x modelled size?

- We adjusted the component GLM predictions by the output from the scoring runs
- Determined a cost per policy prediction as the product of these “enhanced” models
- Ran a Complete Segmentation at cost level
- Discovered a few segments where the cost prediction did not fit well
- Examined the component model fit for those segments (also checked the base GLM)
- The size model is the culprit but this is not clear until we look at the fitted average size weighted by the expected frequency

<table>
<thead>
<tr>
<th>Segment</th>
<th>% Exposure</th>
<th>Freq</th>
<th>Size¹</th>
<th>Cost</th>
<th>Size²</th>
<th>Freq</th>
<th>Size¹</th>
<th>Cost</th>
<th>Size²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.8%</td>
<td>96</td>
<td>99</td>
<td>87</td>
<td>90</td>
<td>99</td>
<td>93</td>
<td>84</td>
<td>85</td>
</tr>
<tr>
<td>2</td>
<td>8.9%</td>
<td>99</td>
<td>97</td>
<td>91</td>
<td>91</td>
<td>100</td>
<td>96</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>Other</td>
<td>78.0%</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>7.2%</td>
<td>103</td>
<td>104</td>
<td>126</td>
<td>122</td>
<td>108</td>
<td>105</td>
<td>134</td>
<td>124</td>
</tr>
</tbody>
</table>

1 – Based on scoring of policies with a claim used in size model
2 – Implied average size based on cost scoring of entire exposure file
Scoring Run against Total Cost Prediction

- Chart shows a scoring run against a CART enhanced total cost prediction
- Model shows good lift and good persistence of signal across the validation data
Using Machine Learning

• Why use a GLM? (vs ensemble prediction)
  – Enables a deep understanding of drivers
  – Forces us to get close to the data and understand it
  – “Easy” to communicate

• There are limitations
  – Linearity
  – Response distribution
  – Independence
  – Practicality
Predictive Model Improvement

• Component model by peril:
  – Use segmentation to guide search for complex interactions
  – Incorporate these findings into the GLM

• Perform segmentation on peril cost prediction
  – Adjust prediction outside the GLM

• Examine total cost prediction
  – Perform segmentation and adjust outside the GLM

• Use scoring to incorporate an additional complex layer which improves predictive power
Other Applications of Machine Learning

Useful segmentation targets include:

- Loss ratio (or COR)
- Joint profitability / sales or profitability / competitive position
- Instability
- Price elasticity
Conclusions

- GLMs are a powerful tool for building predictive models
- There are inherent assumptions and other practical challenges
  - Probabilistic inference to test a hypothesis
- Limitations compromise predictive accuracy
  - “Sub” optimisation
- Machine learning techniques can be used to improve predictive models
  - Using CART is not enough