Network Analytics in Claims Level Predictive Modelling

Marcela Granados, Satraajeet Mukherjee

“R in Insurance” conference
Paris, France
June 8th, 2017
Interaction Between Claims and Analytics

► What drives adverse development?
  ► Adverse development is disproportionately driven by specific types of claims
  ► It can be difficult to quantify the preponderance of factors that drive claims development
  ► Early identification of these claims allows for proactive claims handling and real cost savings

![Diagram illustrating interaction between claims and analytics]

1. **Claims predictive modeling**  
   Application of advanced analytics to identify drivers of unexpected development

2. **Claims triage and mitigation**  
   Implementation of model to predict development and apply loss mitigation strategies

3. **Claims leakage analysis**  
   Analyze unexpected development to identify inefficient claim root cause practices

4. **Process improvement**  
   Incorporate findings into claims and underwriting processes to develop a leading practice environment

Driving cost savings throughout the process
Claims Life Cycle

**WC Claim # 1**

- Information available at claim intake (FNOL)
  - Structured Data – Accident Date, State, Industry, etc.
  - Unstructured Data – claim adjuster notes

- State - High Severity, Litigious
- Occupation - Hotel Luggage Services: High Severity
- Injury Type - Strain: Medium Severity

- Unstructured Data - Strain to lower back by lifting heavy objects

**WC Claim # 2**

- Information not captured in structured data

- State - Medium Severity
- Occupation - Hotel Cleaning Services: Low Severity
- Injury Type - Strain: Medium Severity

- Unstructured Data - Strain to back by slipping and falling on wet floor
NLP is used to clean unstructured data, then network analytics is used to identify predictors.

Data Cleaning steps:
1. Remove stop words (e.g. ‘to’, ‘and’, etc.)
2. Stemming (identify words that have the same meaning/root)

Create network graph:
1. Node size (circle) shows the frequency of words
2. Width of lines between nodes shows frequency of words occurring together

Tutorial

CLM_DESC
Strain to left shoulder while opening....
Slipped and fell on wet flooring....
OV rear ended employees vehicle....
Strain to lower back while....
Strain to back
Struck by housekeeping cart....
Valet Scrapped side of guests car against pole....
Alleges ceiling fan blade fell on his head, causing
Strained left back, while lifting....
Strain to left shoulder while....
Contustion to right wrist from....
Slipped and fell bathroom floor....
Tripped over bedspread, causing....

Inputs for Predictive Models

R Packages for Machine Learning
- randomforest, caret

Raw Unstructured Claims Data
Powerful visual analytics tools such as Geochart and wordcloud can be used to analyze structured and unstructured data to identify the most predictive variables.

Darker Green indicates higher probability of litigation.

Larger words indicate higher frequency of words.
Machine Learning – Model Evolution

Parameter Estimation in Logistic Regression

\[ L(\beta|y) = \prod_{i=1}^{N} \frac{n_i!}{y_i!(n_i-y_i)!} \pi_i^{y_i} (1-\pi_i)^{n_i-y_i} \]

Criteria for Random Forest Modelling Splits

\[ Weighted \ Gini = \sum_i \frac{n_i}{N} (p_i^2 + q_i^2) \]

\[ Entropy = -p \log_2 p - q \log_2 q \]
Use of Machine Learning – Random Forests

- Random forests build upon the concept of asking the classification question to multiple people who think differently, such that the end answer is truly unbiased.

- So instead of relying upon a single decision tree and dataset, the algorithm builds an ensemble of decision trees using bootstrapped versions of the original dataset.
Next Steps

Random Forests are not as good as for regression problem as it does not give precise continuous nature predictions.

Random Forests may over-fit data sets that are particularly noisy.

Current

Future

Model Enhancements

Gradient Boosting

Use Cross Validation

Model Application

Current

Future

Probability of Litigation on a particular claim

Identify Fraud Claims

Probability of Large Claim
Shiny Interface – Network Graph
Shiny Interface – Model Results
Shiny Interface – Claims Adjuster’s tool

<table>
<thead>
<tr>
<th>Claim</th>
<th>48015871</th>
</tr>
</thead>
<tbody>
<tr>
<td>State</td>
<td>MA</td>
</tr>
<tr>
<td>Gender</td>
<td>M</td>
</tr>
<tr>
<td>Predicted Probability</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Showing 1 to 4 of 4 entries

- Prominent Words
  - shoulder
  - strain

Showing 1 to 2 of 2 entries