Delivering Value from Unstructured Data with Machine Learning

Tony Ward, Machine Learning Solutions
Unstructured data and the 80% Rule

LOSS ADJUSTER NOTES

UNDERWRITER NOTES

EMAILS

RISK REPORTS

SOCIAL

WEB PAGES

JOB ADVERTISEMENTS
Motivating Example

- Web Scraped 18,000 Jobs from [www.theactuaryjobs.com](http://www.theactuaryjobs.com)
- Structured Data
  - Sector
  - Experience
- Unstructured Data
  - Job Title
  - Job Description
Global insurance company with Lloyds syndicate is looking to hire a pricing actuary reporting directly into the Group pricing actuary you will have experience of pricing and be looking to lead a small team based in London. The business covers a variety of lines of business including marine, aviation, energy, property and liability. You will work closely with the underwriters, present to senior management and underwriters and own and manage rate adequacy and rate movement MI for international specialty lines of business. Above and beyond your pricing skills you must be able to demonstrate leadership qualities and have excellent communication skills.
Job 2

Our client, an investment consultancy, is looking to hire a research analyst. You will work on investment manager research projects. Working closely with the investment consulting team, you will conduct research on multi-asset fund managers and equity fund managers. Candidates are expected to have experience in researching investment managers and in interest in financial markets and funds. You must be a good team player with good communication skills and be able to present to the investment committee. Our client offers study support (e.g. CFA or actuarial qualifications).
It’s tough

• Can’t read all 18,000 job descriptions
• Delivering insight from this kind of data is hard
• We need a new set of techniques
Today we will

- Apply a machine learning technique called Topic Modelling to the Job Description.
- Visualise this model interactively
- Gain insight into the different types of Actuary (Number 26 will shock you!)
- Test how well this model generalises to new data
How to handle unstructured data

- Turn the unstructured data into a structured tabular form (a Document Term Matrix).

<table>
<thead>
<tr>
<th>Job Title</th>
<th>Analyst</th>
<th>Actuary</th>
<th>Head</th>
<th>International</th>
<th>Pricing</th>
<th>Research</th>
<th>Retail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head of International Pricing</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research Analyst</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Pricing Actuary - Retail</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

- Do this for the “Job Description”
- Now we can Visualise the most frequent terms
Other tricks

- Remove Most Frequent Words
- Remove Very Infrequent Words
- Convert to Lower Case
- Remove Punctuation
- Remove Numbers
- Replace Synonyms
  - Handsome: Hunk, Heart-Throb, Prince Charming
Topic Modelling is..

- Probably my favourite ML technique
What is a topic?

- A topic is a probability distribution
Example

Topic 10

Actuary Analysis Capital Data Economic Emblem Igloo Internal Model Pricing Product Rate Remetrica SAS Solvency Statistical Validation

0% 5% 10% 15% 20% 25% 30%

06 October 2016
Another Example
Mixtures are allowed

Seeking a part qualified or qualified non-life actuary with knowledge of Emblem and SAS to drive innovative project work within our client’s growing pricing team. An outstanding opportunity to develop your career within this market-leader. Please contact us to discuss this vacancy or for an informal discussion regarding your career goals. We are very happy to perform bespoke research on your behalf.
Mixtures are allowed

Seeking a qualified actuary with detailed experience of GI, actuarial & risk modelling software to provide modelling solutions to prospective clients. You will provide actuarial, business and technical input to RFPs and client proposals, based on an in-depth understanding of regulatory requirements and ERM solutions. An unrivalled opportunity to develop the strategic vision of this market-leader. Please contact us to discuss this vacancy or for an informal discussion regarding your career goals. We are very happy to perform bespoke research on your behalf.
## Topic Model Output

<table>
<thead>
<tr>
<th>Term</th>
<th>Topic 1</th>
<th>Topic 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actuary</td>
<td>25%</td>
<td>23%</td>
</tr>
<tr>
<td>Analysis</td>
<td>0%</td>
<td>5%</td>
</tr>
<tr>
<td>Capital</td>
<td>18%</td>
<td>0%</td>
</tr>
<tr>
<td>Data</td>
<td>0%</td>
<td>3%</td>
</tr>
<tr>
<td>Economic</td>
<td>5%</td>
<td>0%</td>
</tr>
<tr>
<td>Emblem</td>
<td>0%</td>
<td>15%</td>
</tr>
<tr>
<td>Igloo</td>
<td>9%</td>
<td>0%</td>
</tr>
<tr>
<td>Internal</td>
<td>9%</td>
<td>0%</td>
</tr>
<tr>
<td>Model</td>
<td>16%</td>
<td>0%</td>
</tr>
<tr>
<td>Pricing</td>
<td>0%</td>
<td>23%</td>
</tr>
<tr>
<td>Product</td>
<td>0%</td>
<td>9%</td>
</tr>
<tr>
<td>Rate</td>
<td>0%</td>
<td>3%</td>
</tr>
<tr>
<td>Remetrica</td>
<td>10%</td>
<td>0%</td>
</tr>
<tr>
<td>SAS</td>
<td>0%</td>
<td>14%</td>
</tr>
<tr>
<td>Solvency</td>
<td>5%</td>
<td>0%</td>
</tr>
<tr>
<td>Statistical</td>
<td>0%</td>
<td>5%</td>
</tr>
<tr>
<td>Validation</td>
<td>3%</td>
<td>0%</td>
</tr>
</tbody>
</table>

*INTERACTIVITY IS NOT OPTIONAL*
Live Demo.....
Validation
Putting this into Production
Application 1: New Perils
Application 2: Underwriter Notes
Application 3: Social Data
Application 4: Over to you....
Thank you!

Tony Ward

www.machinelearningsolutions.co.uk