Institute and Faculty of Actuaries

Maximising the Value of Text Based Insurance Data

Santiago Restrepo L. (BDO)
Dr. Rickard Nyman (Periander)

10 September 2019


actuaries AI algorithms answer applications apply based basics book categories claim computing data
document example experience expert form frequency given information insurance judgment key known
learning likely lot machine mining non-numeric number operations past policies pool present price recent representations result risk
search similarity structured talk techniques term text TF today type underwriters unstructured
used value vector weights words works

AI and new technology

- AI / Data Science / Machine Learning have become buzz words.
- Many of the AI techniques used today are enhancements of concepts that have been around since the 1960's.
- There are a lot of machine learning and AI algorithms that have not yet made their way into actuarial practice and probably never will.
- Insurance has used structured data for decades and it is one of the industries that has been data driven since its inception.
- Unstructured data does not go un-analysed. Just not systematically or in a way that can be objectively measured.

Case-Base Reasoning and expert judgment

CBR

- Paradigm of Artificial Intelligence
- Solve problems by retrieving similar stored cases
- Adapt stored case to fit new need – what is different?
- Support human decision making, aid learning, facilitate access to information

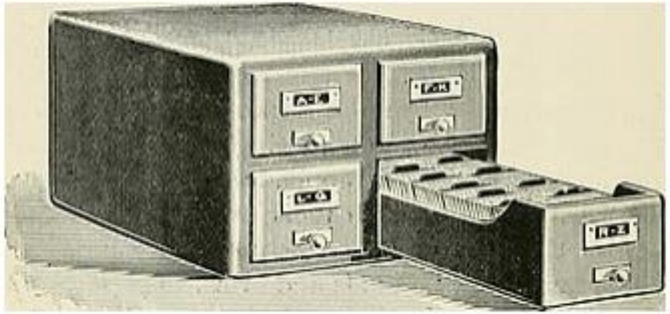
Actuarial Expert Judgment – Reserving example

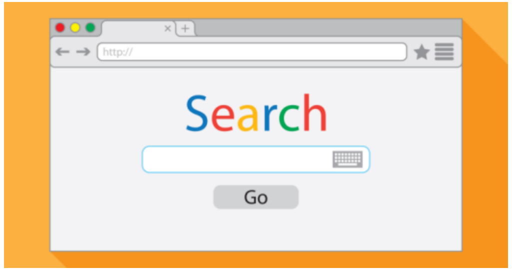
- We observe mix of business changing
- CL development pattern may need to shift for recent cohort of policies
- Retrieve cases of what is similar to current mix
- Apply appropriate pattern for recent cohorts

Institute and Faculty of Actuaries

Pools, tools and mind-set

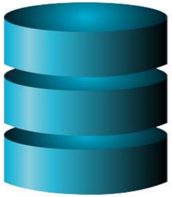
- Let $\{x\}_{i=1..n}$ be a set of n independent and identically distributed r.v.
- Groups of similar risks that are managed together
- Homogeneous Cohorts
- Covariates in a GLM
- etc



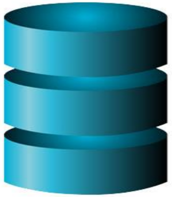


Institute and Faculty of Actuaries

The Policy/Claim metric space

Policy 1

Policy 2

Institute and Faculty of Actuaries

Text Analysis

- The methodologies have become more powerful in the last few years
- One approach, for example, is to look for a pre-determined set of words/phrases in a text
  - For example, in economics, the Economic Policy Uncertainty Index has gained traction, and this is (in part) based on mentions of the word “uncertainty” in various newspapers
  - Another example, **Rickard Nyman** et al., *News and narratives in financial systems: Exploring big data for systemic risk assessment* (Bank of England Working Paper No. 704). 2018.
- Increasingly, the process is to assign a fairly small number of texts into the categories of interest based on the overall impression given by the entire text
- Algorithms learn the words/phrases which classify the texts into the categories successfully

Institute and Faculty of Actuaries

The Vector Space Model

- First mathematical model of text for machine manipulation
- The VSM model first used in the SMART (System for the Mechanical Analysis and Retrieval of Text) Information Retrieval System developed at Cornell in the 1960s
- Represent a document *d* by an index vector $\boldsymbol{x} = (\boldsymbol{x\_1}, \ldots, \boldsymbol{x\_m})$ of dimension m
- A common document representation is known as the n-gram model (segment d into n-tuples of sequences of tokens)
- Typically the order of n-grams is ignored (‘bag of words’ model)

Institute and Faculty of Actuaries

Encoding Schemes

- x can be encoded in a number of ways, common encoding schemes:
- For a vocabulary of all n-grams $\boldsymbol{V}=(\boldsymbol{v\_1},\ldots,\ \boldsymbol{v\_N}\ )$ define $\boldsymbol{x\_i}\in\{\mathbf{0},\mathbf{1}\}$ to be 1 if the n-gram is in *d* and 0 otherwise (Boolean). So N = m
- For vocabulary $\boldsymbol{V}=(\boldsymbol{v\_1},\ldots,\ \boldsymbol{v\_N}\ )$ define $\boldsymbol{x\_i}\in\mathbb{N}$ to be the frequency of the n-gram in d, $\boldsymbol{tf}(\boldsymbol{v\_t})$

Institute and Faculty of Actuaries

Encoding Schemes

- Let K be the number of documents in the database D. The inverse document frequency of an n-gram $\boldsymbol{v\_t}$ is defined as:

$$\boldsymbol{idf}(\boldsymbol{v\_t},\boldsymbol{D})=\boldsymbol{log}\ \boldsymbol{K}/|\{\boldsymbol{d}\in\boldsymbol{D}:\boldsymbol{v\_t}\in\boldsymbol{d}\}|$$

- Term frequency inverse document frequency representation is defined as: $\boldsymbol{tf}(\boldsymbol{v\_t})*\boldsymbol{idf}(\boldsymbol{v\_t},\boldsymbol{D})$. A balanced measure of frequency of an n-gram in a document weighted by its ability to ‘distinguish’ between documents in general in the database D
- Similarity between any two vectors is typically measured by the angle between them, or equivalently cos(angle)

Institute and Faculty of Actuaries

Word embeddings: GloVe

- The approach (known as GloVe) described by Pennington et al. (2014) is widely used to derive word-vector embeddings (the paper already has 6000 citations)
- The training objective of GloVe is to learn word vectors such that their dot product equals the logarithm of the words' probability of co-occurrence. Owing to the fact that the logarithm of a ratio equals the difference of logarithms, this objective associates (the logarithm of) ratios of co-occurrence probabilities with vector differences in the word vector space
- It is built up from around 2 billion words in English from Twitter, Wikipedia, web crawls and can be downloaded from https://nlp.stanford.edu/projects/glove/

Word embeddings: Word2Vec

- A highly influential recent development using neural networks (Mikolov et al. 2013)
- There are two distinct approaches
- Both rely on training neural networks to predict properties of word sequences
- CBOW: Suppose we have a sequence of n words, where n is an odd number. We leave out the middle word and train a neural net to predict what it is from the surrounding ones. Use the weights of the hidden layer as feature vectors for the target word

Institute and Faculty of Actuaries

Word embeddings: Word2Vec

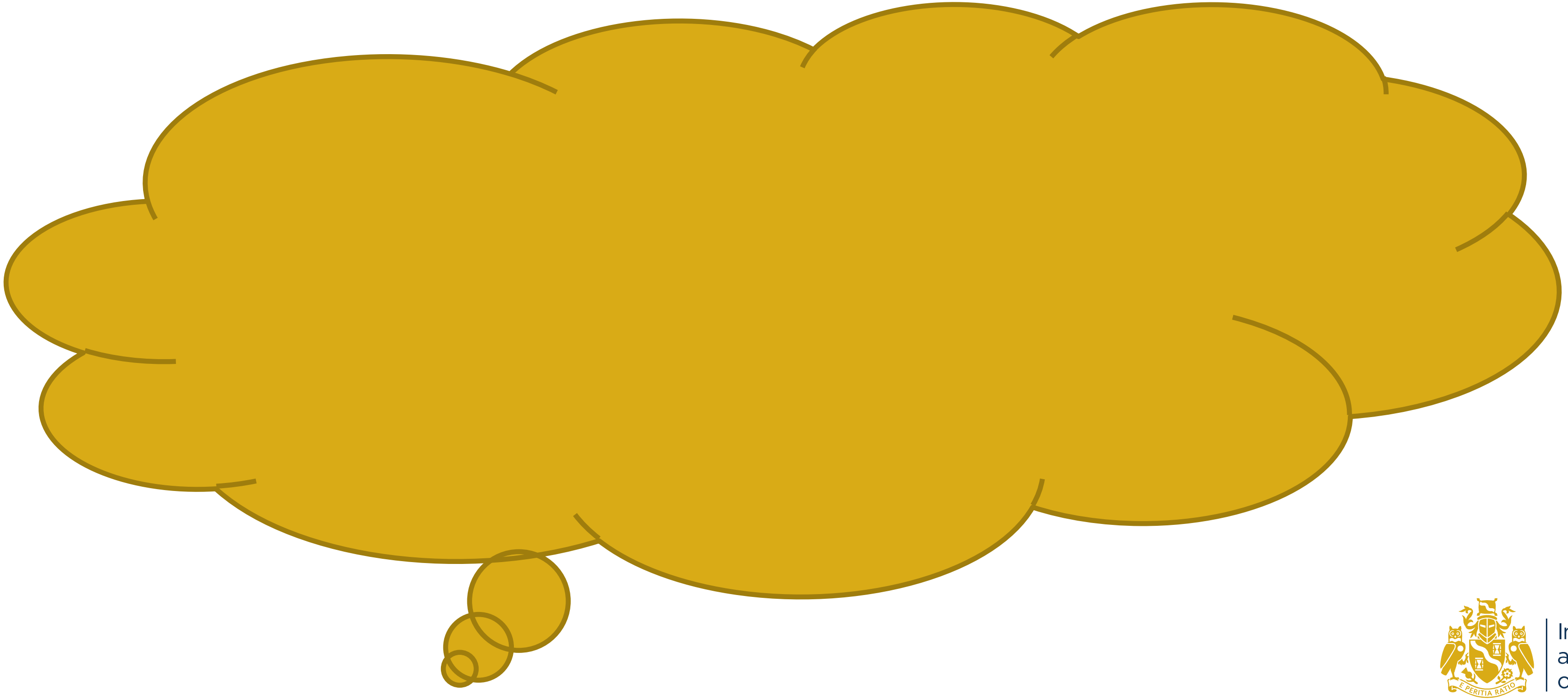
- Skip-Gram: Suppose we have a sequence of n words, where n is an odd number. We leave out the surrounding words and train a neural net to predict what they are from the centre word. Use the weights of the hidden layer as feature vectors for the centre word
- Mikolov et al. show that the relatively simple neural network architectures they propose achieve high quality word-vector representations comparable to more complex networks structures such as Recurrent Neural Nets while being much faster to train on larger data sets

Text analysis – tech difficulties

- A purely technical point is that much of the text held by companies may not be in formats which are ideal for reading into text analysis algorithms
- But this is a solved problem
- The quality of the data is important e.g. a large amount of very bland text is of little use
- Resistance because of a worry that jobs will be displaced

Institute and Faculty of Actuaries

The policy and claim search engine

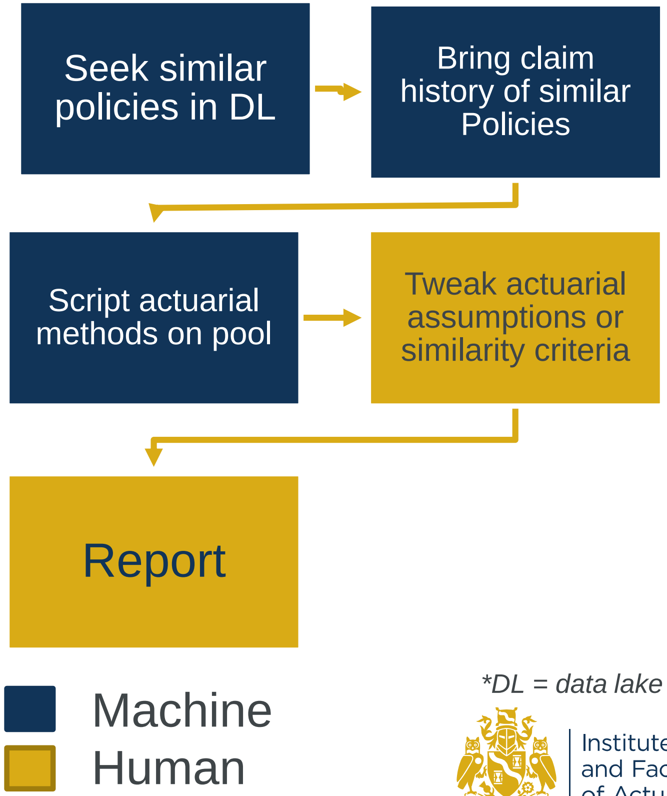


Important aspects

- The method complements actuarial methods and judgment– it does not replace them.
- Uses AI and other non supervised approaches
  - But the results can be validated by inspecting the policies deemed similar
  - The search results can be adjusted to reflect what the human intelligence deems relevant for similarity in the specific context
  - Company specific expert knowledge can be factored into the specification of inter-word relationships
    - Latent features not mentioned explicitly in text can be factored in via a separate classification exercise.
    - This not an off the shelf product.

Benchmark pricing Application

- Upload a few documents – type in the required numerical/categorical inputs
- Talk to the pc about your notes – or upload telephone conversations
- Report could include
  - Expected loss cost – by peril
  - Sensitivity – increasing diameter of similarity
  - Market price of recently quoted/signed policies



*DL = data lake

Institute and Faculty of Actuaries

Expert Judgment Validation

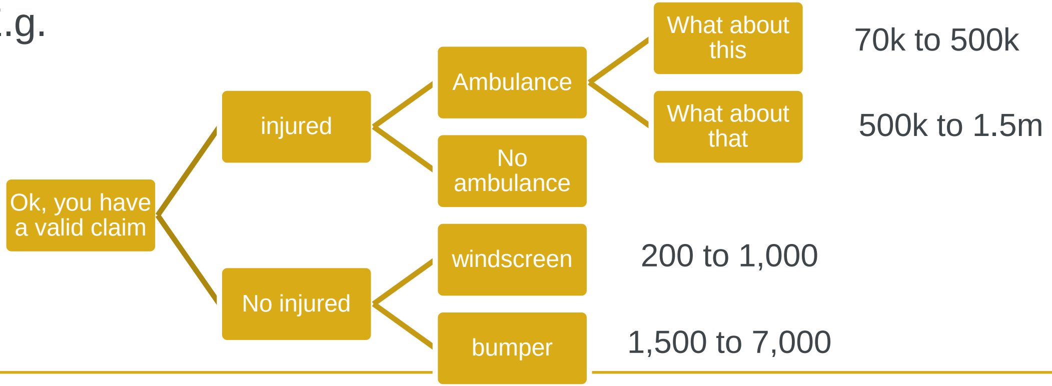
- Underwriter: we have re - underwritten the book and now the loss ratio is lower.. I'd say it is about 15% lower than what you say
- Actuary: How do you know?
- Underwriter : I'm the expert, I've been in this business far longer than you.
- Actuary: can you prove it? How do you identify from the policy documentation these 'low frequency' risks?

Actuary then designs a search for a pool with the described key identification factors and check whether the frequency of such policies is lower than the overall frequency

Institute and Faculty of Actuaries

Claim triage

- Multinomial GLMS have been used for triaging claims
- Predictive strength could be improved with text analysing claim narratives by
  - Finding sets of related words that could lead to higher costs or delayed settlements
- Help in claims adjusting by suggesting interrogation points based on the narrative so far and with the aim of minimising the range of potential outcomes
  - E.g.



Why this matters

- Low interest rate environment – intelligent pricing, finding profitable niches and reducing adjustment expenses will become more important
- Could sound futuristic but
  - Electronic placement will make the necessary information available and easier to e-analyse
  - Most paperwork is already electronic
  - The search and text-crunching technology is out there
  - We already operate with this mind-set outside of our work place!

Institute and Faculty of Actuaries

Questions

Comments

The views expressed in this [publication/presentation] are those of invited contributors and not necessarily those of the IFoA. The IFoA do not endorse any of the views stated, nor any claims or representations made in this [publication/presentation] and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this [publication/presentation].

The information and expressions of opinion contained in this publication are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this [publication/presentation] be reproduced without the written permission of the IFoA [*or authors, in the case of non-IFoA research*].

Institute and Faculty of Actuaries