Non-Life Insurance Pricing using R

Deploying advanced analytics in the Insurance industry

R in Insurance conference
Cass Business School 15 July 2012

64 Squares and CYBAEA
Suresh Gangam (suresh@64sqs.com)
Allan Engelhardt (Allan.Engelhardt@cybaea.net)

Background

I am interested in how analytics fits into…

…all the other things we do in a business or other organization.

In other words, how do we turn mountains of data…

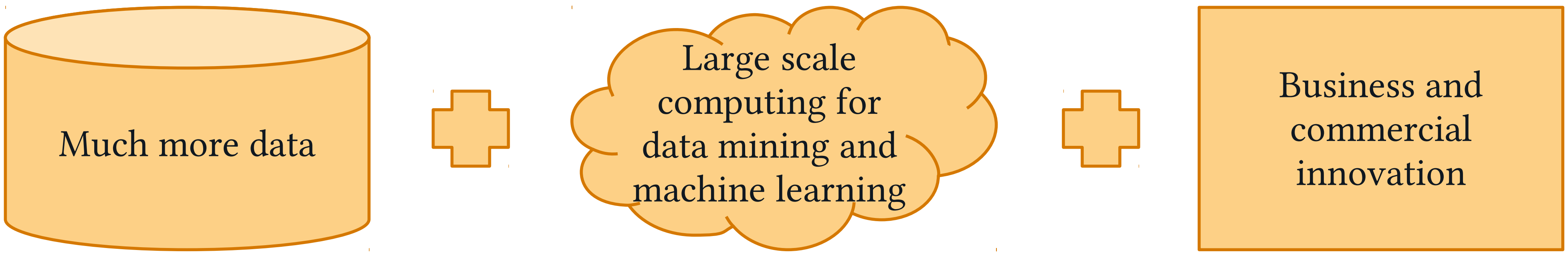
…into mountains of money (metrics of successful actions).

Now that the technology is finally ready for the enterprise…

…we need to make sure the enterprise is ready for the technology.

Advanced analytics and ‘Big Data’ is disrupting most industries

- Step change in the volume and types of data that are now used for analytics
- Substantial change in our ability to leverage large data
  - Cloud computing has changed the economics of large scale computing
  - Latest data mining and machine learning techniques are now readily and economically available through tools like R which has massive vendor support
    - Standalone or embedded in the enterprise technology stack including data warehouse, message bus, and dashboards – massive support from enterprise vendors like Oracle, SAP, Tibco, Teradata, EMC, and many more on a rapidly growing list



Disruptive innovation through advanced analytics

Advanced analytics and 'Big Data' is disrupting most industries

- Step change in the volume and types of data that are now used for analytics
- Substantial change in our ability to leverage large data
  - Cloud computing has changed the economics of large scale computing
  - Latest data mining and machine learning techniques are now readily and economically available through tools like R with massive vendor support
- Commercially exploiting the opportunities of Big Data is essentially understood and proven
  - Strategy, organization, process, proposition

Innovative insurers are at an advantage

- Insurers are under pressure to improve their analytical capabilities by using more data and advanced modelling techniques
  - For example, Tesco Insurance has access to Clubcard holders' detailed shopping history at the time they make an underwriting decision. They can decide your car insurance premium knowing how much alcohol you buy. They have a competitive advantage by using more data and large scale data mining and machine learning techniques.

Innovative insurers are at an advantage

- Insurers are under pressure to improve their analytical capabilities by using more data and advanced modelling techniques
  - For example, Tesco Insurance has access to Clubcard holders' detailed shopping history at the time they make an underwriting decision. They can decide your car insurance premium knowing how much alcohol you buy. They have a competitive advantage by using more data and large scale data mining and machine learning techniques.
- Traditional insurance companies need to respond to innovative new challengers in the market or risk going the way of the small supermarket and independent book store (anybody remember them?)
  - Like in so many other industries, the combination of availability of large amounts of data with the processing power to analyse it and the business innovation it promotes is a *disruptive innovation* (in the Clayton Christensen sense) in Insurance.

Innovative insurers are at an advantage

- Insurers are under pressure to improve their analytical capabilities by using more data and advanced modelling techniques
  - For example, Tesco Insurance has access to Clubcard holders' detailed shopping history at the time they make an underwriting decision. They can decide your car insurance premium knowing how much alcohol you buy. They have a competitive advantage by using more data and large scale data mining and machine learning techniques.
- Traditional insurance companies need to respond to innovative new challengers in the market or risk going the way of the small supermarket and independent book store (anybody remember them?)
  - Like in so many other industries, the combination of availability of large amounts of data with the processing power to analyse it and the business innovation it promotes is a *disruptive innovation* (in the Clayton Christensen sense) in Insurance.
- Fortunately, advanced analytical capabilities are now available for the taking.
  - The latest machine learning techniques are easy to access through modern analytical tools like R and large scale computing is readily available through private and public clouds.
  - It has been done before and it can be done incrementally.

Incremental approach to advanced analytics capabilities

- Rather uniquely, the insurance industry is able to respond to this challenge by incremental improvements
  - The value changes that are needed are ones that are already in progress and accepted in principle: Products to People
    - Primarily: to move from managing books of policies to managing customers. This change from products to people is one that many industries (example: mobile telcos) have done

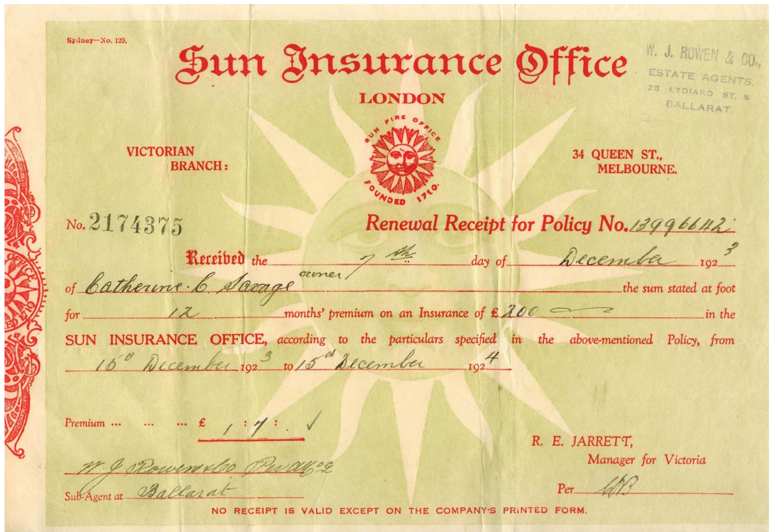


Image by Laineys Repetoire on Flickr

Image modified from original by Dominic's Pics on Flickr

Incremental approach to advanced analytics capabilities

- Rather uniquely, the insurance industry is able to respond to this challenge by incremental improvements
  - The value changes that are needed are ones that are already in progress and accepted in principle: Products to People
- You do not have to "boil the ocean" to get started. With the right commercial focus you will deliver value immediately even from small, easy-to-implement changes, and you will have started your journey.
- You do not need to abandon well-established practices but can supplement them and validate them against best-in-class processes to deliver commercial value right now

Case Study

Very brief introduction to non-life insurance pricing

- The question we are considering is *tariff analysis*: how much to charge an individual policyholder within an insurance portfolio (given an overall premium level for the book).
- The usual approach is to model using generalized linear models (GLM) a number of *key ratios* as dependent on a set of *rating factors*.
  - For personal lines the key ratios are often claim frequency and claim severity (cost per claim) while for commercial lines we may consider the loss ratio (claim costs per earned premium).
  - Rating factors are <u>grouped into classes</u> (i.e. factor variables) and may include
    - Information about policyholder: age, gender, line of business, etc.
    - Information about the insured risk: age and model of car, type of building, etc.
    - Geographic and demographic information: population density, income levels, etc.
  - A given value for the rating factors is called a *tariff cell*.

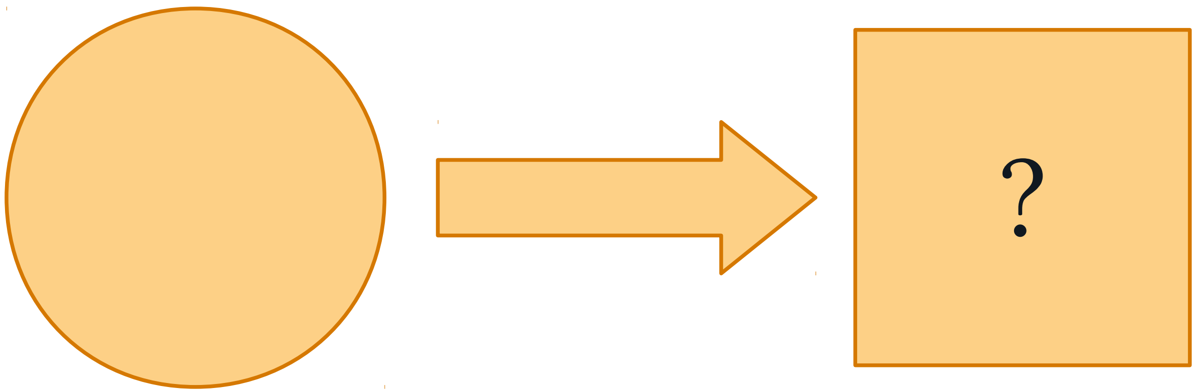
Very brief introduction to non-life insurance pricing

- We assume
  - Policy independence: claims are independent across policies.
    - No catastrophes, no collisions, …. (Reinsurance can help with the first.)
  - Time independence: claims over different times are independent.
    - The world is static. (Factor out inflation and similar by considering price ratios.)
  - Homogeneity: claims are uniform within tariff cells.
    - Use bonus/malus systems and experience ratings (of companies) to cope with non-homogeneity.

Exposure w	Response X	Key ratio X/w
Duration	Number of claims	Claim frequency
Duration	Claim cost	Pure premium
Number of claims	Claim cost	Claim severity
Earned premium	Claim cost	Loss ratio
Number of claims	Number of large claims	Proportion of large claims

Squaring the circle for analytics

- Typical management requirement for analytics:
  1. Must be an interpretable model
  2. Must take into account the latest analytical, statistical, and industry approaches
- We will show one way to square this circle with an example from pricing, but applications are wider



Case Study:
Validating and extending pricing model

Key Challenge:

- A US based home insurer was facing profitability challenges. The pricing model was inadequate to price all risks appropriately. There were also regulatory constraints around tweaking the pricing model. It was imperative for the insurance carrier to improve profitability

Primary Objectives:

- Evaluate inadequacies in the then current pricing model to identify policies priced far below adequate levels
- Develop a comprehensive strategy to take such policies off the book

Approach:

- Visualized premium and losses against each pricing factor and identified factors where the pricing was inadequate
- Used multiple machine learning models to develop a superior pricing model to identify heavily underpriced policies
- Developed an initial strategy to shelve (not renew) these policies over time

Key Outcomes:

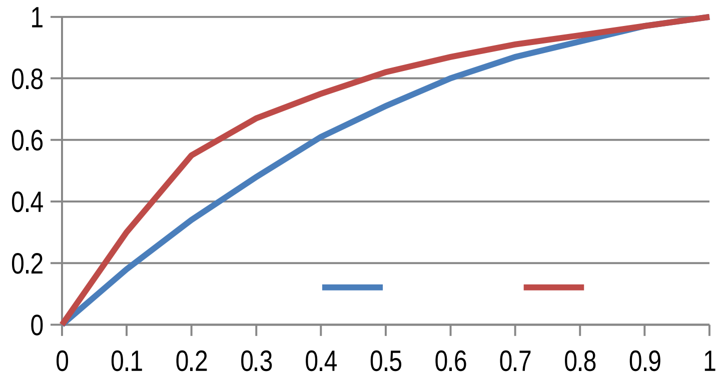
- Insurance carrier was able to address the highly unprofitable policies and improve the profitability to adequate levels
- **The 5% most mispriced policies contributed 14% of the loss ratio**

Step 1:
Explored multiple modern machine learning techniques (GBM, RF, NN, SVM, & more)

Step 2:
Selected GBM and RF for final ensemble model

Step 3:
Test and validate ensemble to show significantly improved model outcomes

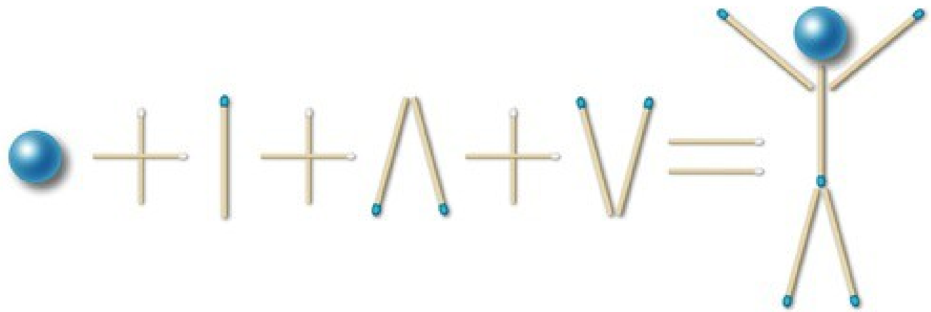
Cumulative % losses for current and new model



Summary

Our approach in summary

1. **Supplement the existing** model approach with more modern techniques
   - Restrict the validity domain of the classical model
   - Create new variables inspired by new model that extends the validity of the old
   - Easy to extend model to consider continuous rating factors, longitudinal data, etc.
     - Though you may need to reshape your input data



Our approach in summary

2. This enables **incremental business change**
    - Risk we do not understand and therefore will not insure
    - Understanding complex risk
        - Creating new variables for GLM
        - Consider GAM, GLMM, and beyond

Our approach in summary

3. Use as a **stepping stone** to a more data-driven enterprise ("Big Data") - *incremental, measurable business results at each step*
   - Keep models reproducible and refreshed by using an appropriate language (instead of point-and-click) on a suitable infrastructure (cloud)
   - Establish the processes and teams around regular model refresh ("model factory")
   - Extend the models to customer view and more
   - Tie in channel performance and sales/marketing campaigns
   - *Keep showing value at each step*

What we gained from R

- Robust process that is almost trivial to extend to different modelling approaches
  - Thank you Max Kuhn and ‘caret’ package

What we gained from R

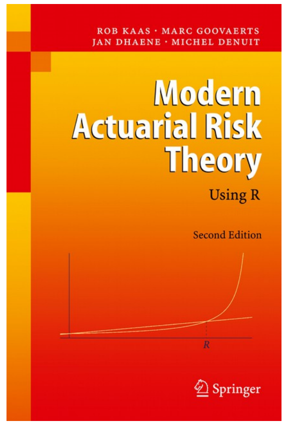
- Robust process that is almost trivial to extend to different modelling approaches
  - Thank you Max Kuhn and 'caret' package
- Perfectly reproducible models by creating and saving virtual image
  - Scriptable language is a key
  - Open source helps

What we gained from R

- Robust process that is almost trivial to extend to different modelling approaches
  - Thank you Max Kuhn and 'caret' package
- Perfectly reproducible models by creating and saving virtual image
  - Scriptable language is a key
  - Open source helps
- A general language and tool that applies across the enterprise and has massive enterprise vendor support
  - Oracle, SAP, Tibco, Teradata, Revolutions, and many more

What we gained from R

- Robust process that is almost trivial to extend to different modelling approaches
  - Thank you Max Kuhn and 'caret' package
- Perfectly reproducible models by creating and saving virtual image
  - Scriptable language is a key
  - Open source helps
- A general language and tool that applies across the enterprise and has massive enterprise vendor support
  - Oracle, SAP, Tibco, Teradata, Revolutions, and many more
- An actively developed tool set with a lively and supportive community
  - Lots of books (search Amazon et al)
    - Look for: Computational Actuarial Science with R (Arthur Charpentier, editor)
  - Conferences (general, finance, insurance, and more)
  - People! (R is what students learn these days)
  - Commercial training



Our approach in summary

1. Supplement the existing model

2. Aim for incremental business change

3. Use incremental change as stepping stones, delivering value at each step

Thank you!
I hope this was useful.
Questions? Comments?

Allan.Engelhardt@cybaea.net
www.cybaea.net

About us

- CYBAEA is the “more than analytics” company that not only delivers the insights from your data but also helps you execute on that knowledge to drive the results that make a difference. CYBAEA is based in London with an international network of associates; visit them at www.cybaea.net.
- 64Squares is a team of data scientists and management consultants working across Asia, Europe and Americas. 64Squares’ mission is to be the world's best provider of advanced machine learning solutions; visit them at www.64sqs.com.
- Together we provide our clients tremendous advantage
  - The best combination of local and remote resources for effective engagement
  - Razor-sharp focus on delivering commercial value, fast through quick projects with measurable value
  - Flexible engagement models including projects, as-a-service, and interim management
  - World-class people, processes, and technologies.