64 squares

CYBAEA
Make money from your data

Non-Life Insurance Pricing using R

Deploying advanced analytics in the Insurance industry

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

Background

I am interested in how analytics fits into…

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

Now that the technology is (finally) enterprise-ready…

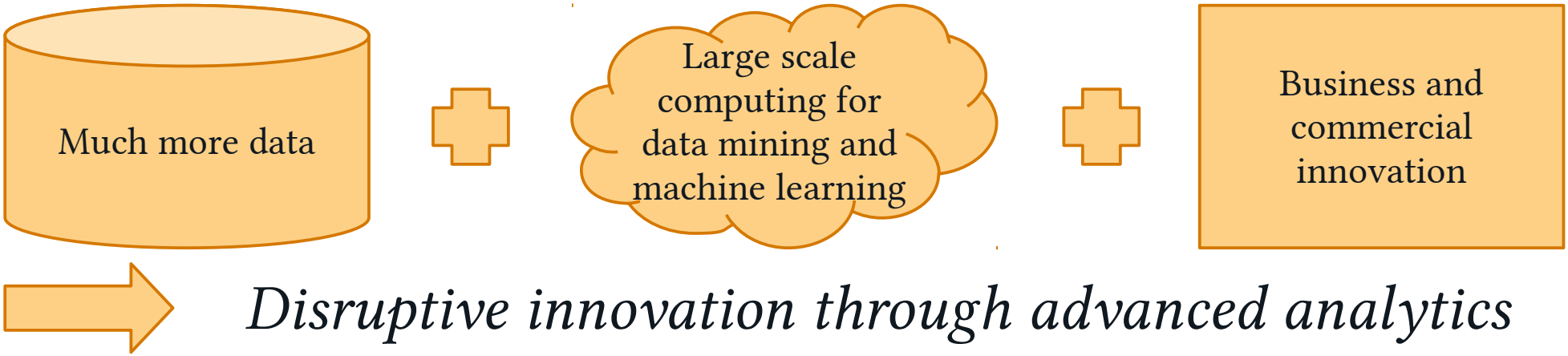
…all the other things we do in a business or other organization.

…into money (metrics of successful actions)

…we need to make sure that the business is able to change and drive change

Some changes in the insurance industry

- Insurance is facing disruption from 'Big Data' like most other industries



Some changes in the insurance industry

- Insurance is facing disruption from 'Big Data' like most other industries
- At the same time beginning the transition from Products to People

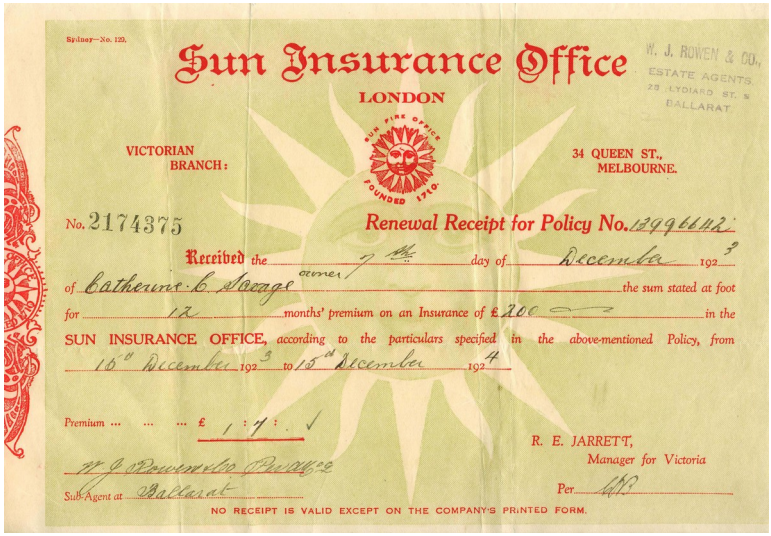

Sun Insurance Office
LONDON
W. J. ROWEN & CO., ESTATE AGENTS, 23 LYDIARD ST. S BALLARAT
VICTORIAN BRANCH:
34 QUEEN ST., MELBOURNE.
No. 2174375
Renewal Receipt for Policy No. 12996442
Received the 7th day of December 1923
of Catherine L. Savage owner
the sum stated at foot
for 12 months' premium on an Insurance of £200 in the
SUN INSURANCE OFFICE, according to the particulars specified in the above-mentioned Policy, from
16th December 1923 to 15th December 1924
Premium ... ... £ 1 : 7 : ✓
R. E. JARRETT, Manager for Victoria
Per
Sub-Agent at Ballarat
NO RECEIPT IS VALID EXCEPT ON THE COMPANY'S PRINTED FORM.

Image by Laineys Repetoire on Flickr

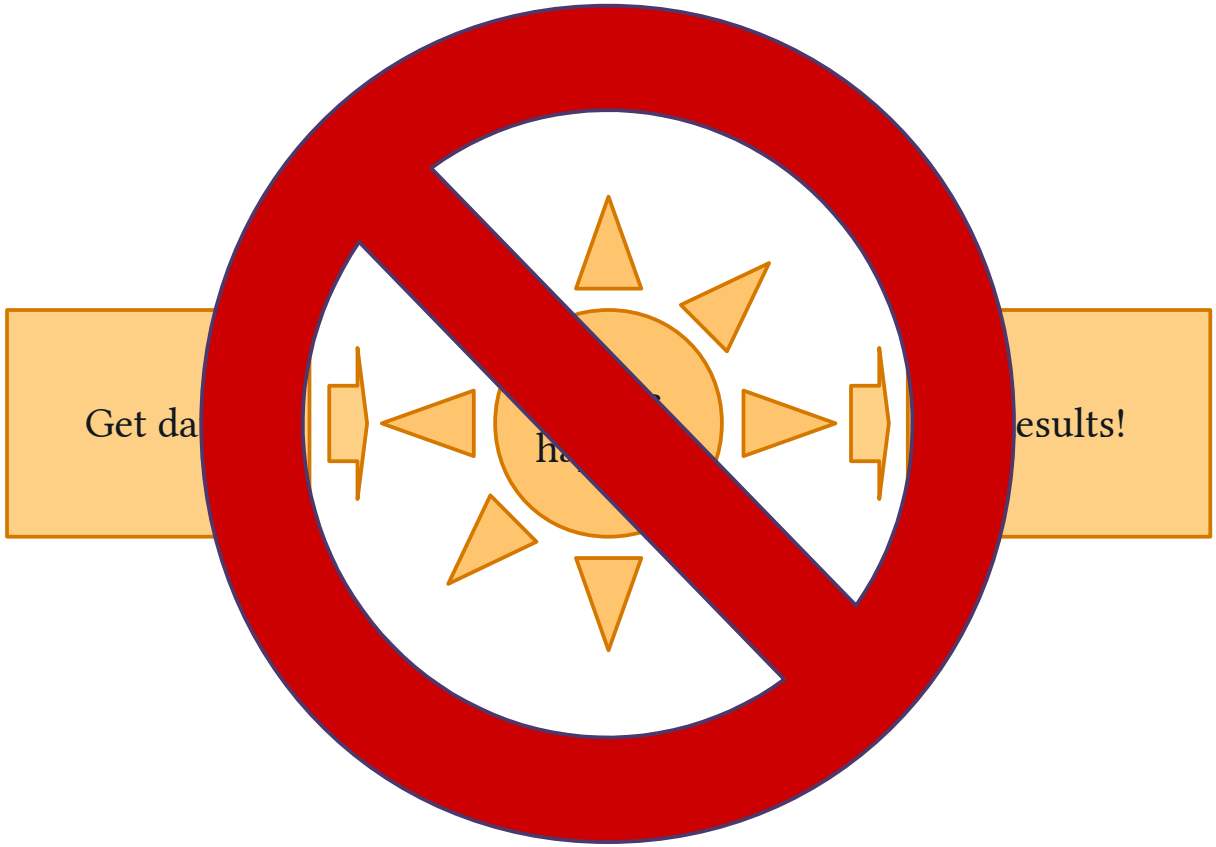
Image modified from original by Dominic's Pics on Flickr

Some changes in the insurance industry

- Insurance is facing disruption from 'Big Data' like most other industries
- At the same time beginning the transition from Products to People
- Analytics capabilities are specialized and compartmentalised
- Regulator is demanding rigorous analytical processes

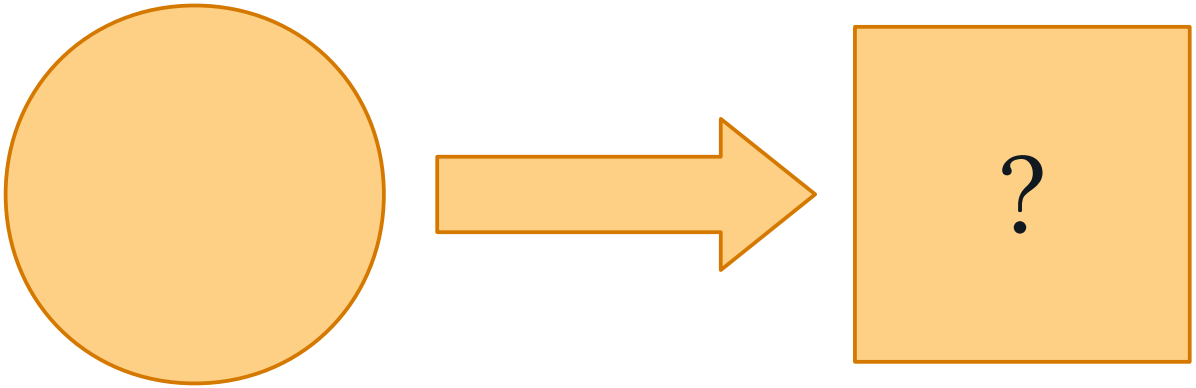
Some changes in the insurance industry

- Insurance is facing disruption from 'Big Data' like most other industries
- At the same time beginning the transition from Products to People
- Analytics capabilities are specialized and compartmentalised
- Regulator is demanding rigorous analytical processes



Some changes in the insurance industry

- Insurance is facing disruption from 'Big Data' like most other industries
- At the same time beginning the transition from Products to People
- Analytics capabilities are specialized and compartmentalised
- Regulator is demanding rigorous analytical processes
- Typical management requirement for analytics
  - Must be an interpretable model
  - Must take into account the latest analytical, statistical, and industry approaches
- We will show how to square this circle with an example from pricing, but applications are wider



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 grouped into classes (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*.

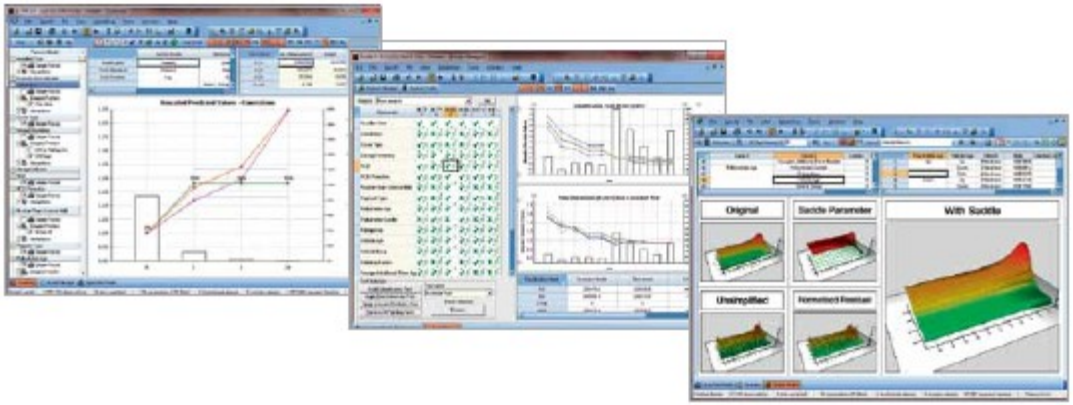
Very brief introduction to non-life insurance pricing

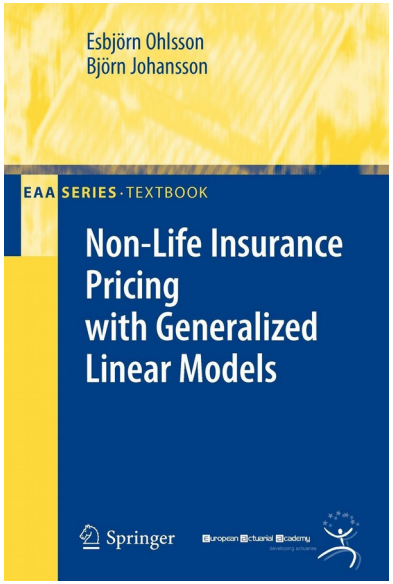
- We assume
  - Policy independence: claims are independent across policies.
    - No catastrophes, no collisions, …. Reinsurance can help with the first.
  - Time independence: claims over different times are independent.
    - The world is static. Factor out inflation and similar by considering price 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

Current implementations

- Despite GLM being a standard statistical theory and readily available in general tools "such as SAS, GLIM, R, or GenStat", most insurance companies rely on proprietary "specialized software provided by major consulting firms".
  - Spreadsheets are also widely used, especially for data preparation but also sometimes directly for analysis, with or without a plug-in.
- Tools are usually applied in point-and-click mode





Some challenges

- Not integrated with anything else and yet pricing is one key component of the overall data usage landscape
  - Moving from managing Products to People means more data sources, more data types, and using different analytical approaches
  - 'Big Data' challenges demands integrated analytics across enterprise data and including external data
- Difficult to extend the modelling approaches
  - Only does one thing
  - What about the "take into account the latest … approaches" requirement?
- Difficult to ensure reproducible results
  - There are some support in commercial tools, but if it is point and click with Excel front and back then not easy

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:

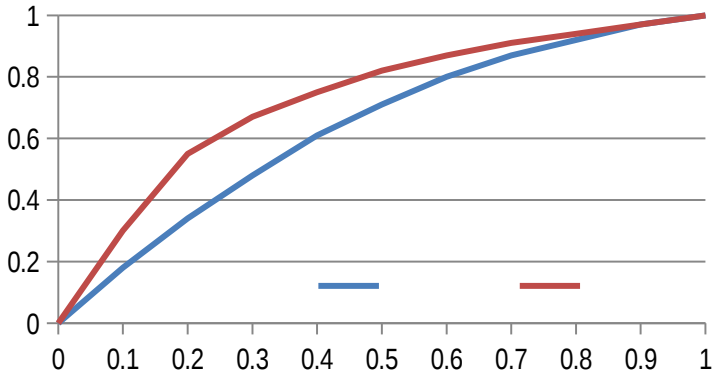
- Insurance carrier was able to address the highly unprofitable policies and improve the profitability to adequate levels
- **The 5% most mispriced policies contributed 14% of the loss ratio**

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



Sketch of R code

```
library("caret")
library("doParallel")                          # or: doMC, doMPI, doRedis, doSNOW
registerDoParallel()

load(file = "Pricing1.RData")

model.1 <- train(
    data, Y, method = "glm",
    trControl = trainControl(method = "cv", number = 5))
model.2 <- train(
    data, Y, method = "gbm",
    distribution = "gaussian", n.minobsinnode = 20000,     # gbm parameters
    tuneGrid = expand.grid(.interaction.depth = c(1,3,5), # Different search
                           .n.trees = c(100, 200, 400),    # to optimize the model
                           .shrinkage = c(1e-1, 5e-2)),    # parameters
    trControl = trainControl(method = "cv", number = 5))
model.3 <- train(
    data, Y, method = "rf", nodesize = 20000,
    tuneLength = 3L,
    trControl = trainControl(method = "cv", number = 5))

print(model.2); plot(model.2); histogram(model.2); # predict(model.2, ...)
```

What we gained from R

- Robust process that is almost trivial to extend to different modelling approaches
  - Thank you Max Kuhn and `caret`
- Perfectly reproducible models by creating and saving virtual image
  - Scriptable language is a key
  - Open source is probably essential for this approach
- A general language and tool that applies across the enterprise
  - Archive data warehouse (Greenplum, Oracle Data Appliance, Teradata Aster, ...)
  - Real-time data store (SAP HANA)
  - Enterprise data bus (supposedly coming at some point)
  - "Model factory" (R, Revolutions, ..., or in the data warehouse)
  - Dashboards (Tibco Spotfire, Shiny), and reports (R)
  - Interactive analytics (R, Revolutions, R Studio, SPSS, ...)
  - Real-time decisioning (*uhm...*, sort of—but not completely there...)

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 and longitudinal data
     - 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")
   - 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: per customer profitability and pricing (and other activities) impact on loyalty & deploy customer centric models widely
   - Tie in channel performance and sales/marketing campaigns
   - And you are nearly there: Keep showing value at each incremental step

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