Development of an application fraud scorecard: A case study of design

Ross Gayler

Baycorp Advantage

This is not a fraud talk*

(* with apologies to Magritte)

Baycorp Advantage

About a Box

SCORECARD CARPENTRY & THE SCORER AS ARTISAN

Baycorp Advantage

What is the most important property?

A box is a composite object

Components are composed to yield the desired box

- 23 separate pieces of wood in this box

Composition vastly increases the space of realisable boxes

- Trinket box
- Bread box
- Sea chest

Function follows form*

(*with apologies)

What goes into making a box? (1)

Structural components (nouns)

- Wood
- Glue
- Wax

Transformational processes (verbs)

- Select
- Cut
- Sand
- Align
- Glue
- …

What goes into making a box? (2)

Design

- What functional properties should the box have?
- What should the composite object (the box) look like?

Realisation

- Choose structural components
- Choose transformational processes
- Compose the transformational processes
- Applying the composite transformational process to the structural components yields the composite object

Design is constrained by the feasible realisations (causal loop)

- Consider design and realisation together

Where is the value added?

Structural components

- Off the shelf (in the store shed)

Transformational processes

- Off the shelf (in the education)

Design & Realisation

- This is the means of expression of the skill of the artisan
- (Decompose the variance of box space into Component, Transformation and Design&Realisation components)

What is the analogy with credit scoring?

Structural components

- Data sets
- Mathematical objects (scorecards, transformations, etc.)

Transformational processes

- Data fitting (regression, calibration, etc.)
- Plumbing (integration of profit component models)

Design & realisation

- This is the means of expression of the skill of the scorer

What is the point?

What gets studied and presented at conferences?

- Structural components
  - Comparison of data sources (e.g. transaction vs summary)
- Transformational processes
  - Comparison of techniques (e.g. linear vs logistic regression)
- Design & realisation
  - Not really (it is implicit)
  - Discussion of product rather than process
  - Assuming skill of the scorer as a given

Where is the value added?

Case study of fraud scorecard development

Look at the reasoning behind the design choices

Shortcomings as a case study

- Limited ability to reveal details because of confidentiality
- No guarantee of optimality of design choices
- Only one alternative design
- No access to design reasoning behind alternative design
- Limited follow up and comparison of performance of designs

Business problem definition

Consumer store card (only useable at a retail chain)

Most applications are received in-store

In-store applications are processed immediately

Some applications are fraudulent, and lead to losses if accepted

Fraudulent applications often use a false identity
(They are not who they claim to be)

Automated application processing can not absolutely confirm identity

Manual checking (e.g. call the employer) detects some frauds

Manual checking adds cost and slows turn-around

Aim to target manual checking at applications most likely to be fraud

Baycorp Advantage

Design goals

Adequate predictive power

- SLA's for approval time & referral rate
  Manual review for fraud increases the referral rate
  Aim for minimum referral rate with adequate fraud hit rate

Durability of predictive power

- Fraud models are known to deteriorate rapidly

Rapid development and implementation

Rapid operational execution

Controllable

Maintainable by current staff

Hard constraints on the design

No programming changes to the application processing system

Minimum logically necessary change in business process

< 30 second approval time (including bureau call)

Only use existing data sources:

- Application form / Customer database / Credit bureau

Operational in 6 weeks (Seasonal variation in fraud rate)

- Obtain development data
- Develop the targeting model
- Implement model in application processing system
- Test the implementation

Design issues
(Categorisation of design problems)

Compatibility of the application processing system and the analytics

- Data mapping between system and analytics
- Form of models

Small number (and low rate) of fraud outcomes

- Fraud is underidentified / Data only available from a limited period
- Tens of thousands of non-frauds / ~400 frauds
- Issues of small and imbalanced samples

Reactivity of fraudsters

- They react to changes in the lending process

Baycorp Advantage

Design principles
(Categorisation of solution features)

Fit the analytics to the processing system

- Use data exactly as it exists in the APS
- Produce models of the form that the APS can implement

Reduce the effective number of predictors (effective vs nominal)

- Include more predictors for better continuity, but
- Model them in a way that uses fewer degrees of freedom

Emphasise cross-validation of models

- Don't fit everything on the one data set

Emphasise predictors that are harder to fake or work for the lender

- Take advantage of the flat-maximum effect

Specific design decisions

Implement the model as standard scorecards

Rapid implementation into existing APS

“Standard” scorecard form

- Not too many predictors
- Limited interaction terms
- Continuous predictors treated as discrete ranges

Possible to develop non-standard models and approximate them as standard scorecards for implementation

- Model continuous variables as continuous then discretise
- Model decision tree as interaction terms

Implement the decision-making using standard referral logic

Generate a score

Compare the score to a threshold

Applications scoring over the threshold are referred for verification

Piggy backs the existing referral logic

Extract the development data from the operational APS

Data in the data warehouse may have been transformed

- Data cleaning
- Encoding of missing values

Guarantees that development is based on the data exactly as it will be seen in the APS

Development data able to be injected into APS to test the implementation of the models in the APS

Segmentation based on data availability blocks (1)

Segmented models are standard and equivalent to interaction terms

Reasonable to include interaction terms but don't want to search over the space of possible interactions

- Overfitting and small data set problems
- Interactions may be more volatile over time
- Constrained time to search

Automated segmentation search generally doesn't ask the right question

- Homogeneity of intercept instead of homogeneity of regression coefficients

Segmentation based on data availability blocks (2)

Predictors are from multiple sources: application form; credit bureau; customer database

Individual predictors may be missing (e.g. field on application form left blank)

Blocks of predictors may be missing (e.g. applicant may not be listed on bureau)

Regression coefficients necessarily vary over availability blocks

- Zero coefficients in segment where missing
- Relative contribution varies as number of predictors varies

Segmented on data availability

Segmentation based on data availability blocks (3)

All applications have most of the application form data

Applicant may be Known to Lender or New to Lender

Applicant may be Known to Bureau or New to Bureau

2 x 2 (Known to Lender x Known to Bureau)

Three segments used:

- Known to Lender & Known to Bureau
- New to Lender & Known to Bureau
- New to Lender & New to Bureau

Known to Lender & new to Bureau is very rare

(pooled with Known to Lender & Known to Bureau)

Baycorp Advantage

Pool segments for early stage development

Not enough data for development of disjoint segment models

- e.g. Known to Lender & Known to Bureau has low fraud rate

Will use multi-stage modelling

- Later stage models are developed by segment
- Early stage models deal with raw predictor variables
- Early stage models based on segments pooled where possible
  - Early stage model using customer database predictors only is based on Known to Lender & Known to Bureau segment
  - Early stage model using bureau predictors only is based on pooled segments: Known to Lender & Known to Bureau and New to Lender & Known to Bureau

Baycorp Advantage

Use more predictors
Use continuous predictors

Aim for a smoothly distributed, continuous score

This allows better control of the referral rate

Must be done in such a way as to avoid overfitting and small data set problems

Weight of evidence coding of predictors

Weight of Evidence is log-odds of a subset relative to the population

Standard practice is to code discrete predictor levels with indicators

One regression degree of freedom per indicator variable (level)

Regression determines the relative contribution of each indicator

Logistic regression coefficients can be interpreted as WOE

Calculate the WOE for each discrete level of the predictor

Use the set of WOEs as a transformation of the predictor

Use the WOE transformed variable as a continuous predictor

Only one regression degree of freedom per predictor

Smooth the WOE for low counts

Multi-stage predictor models to reduce predictor degrees of freedom (1)

Divide predictors into 7 topic groups that are available together

- e.g. address-related, financials, existing customer data

For each topic group build a regression model from the WOE transformed predictors based on the relevant pooled segments.

These models are used to generate 7 subscale scores that are the predictors used in the later-stage models

Subscale models will be applied to all cases, so need to deal with the situation of the predictors being missing.

- Missing score is a constant (log-odds of the missing subset)

Division into subscales reduces the predictor df per regression

Multi-stage predictor models to reduce predictor degrees of freedom (2)

Separate later-stage models for each segment

Predictors are the 7 subscale scores

- Very few predictor df

No subscale score is ever missing

Relative pattern of raw predictors is frozen in the subscale models

The later stage models only alter the relative contribution of the subscales

(WOE coding does the same thing at the lower level)

Emphasise subscale predictors according to pragmatics

Used multi-stage approach within each subscale regression model

Manually selected the ordering of the predictors

Earlier predictors are emphasised

Emphasise predictors that are harder to fake

- Gender of applicant at point of sale
- Bureau inquiries > 12 months ago

Emphasise predictors that work in the lender's favour

- Applicant income

Interaction tree as the final stage

Try to model some interactions (to maximise prediction)

Minimise the interaction variance (because may be less stable over time and may be harder to maintain/control)

Model main effects and simple interactions first (subscale and segment models

Fit and calibrate the regression models then calculate the residuals

Model the residuals using a regression tree

Construct an indicator variable for every node (internal and terminal) of the regression tree

Make those indicators the predictors of the final stage model (use stepwise to select the predictors to keep)

Only marginally improved the model

Fit multiple stages on different data sets to avoid overfitting

Data partitioned into 4 subsets of observations (3 for development, 1 for validation)

Each stage of a multi-stage model should be developed on a separate date set

Unfortunately not enough data for this so cycled through data sets (except the validation set)

The same principle applies to the WOE calculation

- Calculate WOE transform on data set A
- Apply WOE transform to data set B and fit regression model

Reduce the multi-stage models to standard form

The models are linear operators and can be composed and reduced

(The subscale models are multiplied by the appropriate segment model regression coefficient)

Any continuous variables can be divided into arbitrarily fine levels

The end result looks like an ordinary scorecard that can be implemented in the APS

Summary of modelling

Data extracted from operational APS

Raw predictors are WOE transformed

Predictors segmented by topic and availability to fit subscale models

Observations segmented by block availability of predictors

Models from subscale scores fitted separately by segment

Multi-stage modelling used throughout

Specific predictors emphasised based on pragmatics

Separate data sets for each stage (to the extent possible)

Main effects emphasised relative to general interactions

Final model transformed to standard form

Results

This model

- Nominal predictor df of 3 models: ~400
- Effective predictor df: < ~60

Alternative “standard practice” model

- Nominal predictor df of 1 model: ~100

Performance

- Equivalent predictive power at development
- Alternative model significantly worse after a few months
- This model’s predictive power unaltered after a few months
- This model still in use > 5 years after implementation!

Relevance of the case study to the meta-problem

There is a generalisation gradient:

- The specific fraud model
- The specific process used to realise the fraud model
- The pattern of reasoning used to arrive at design and realisation

Credit Scoring & Credit Control IX
7 - 9 September 2005
Credit Research Centre
The University of Edinburgh
Management School

Baycorp Advantage