Model selection and variable aggregation of Australian hospital data

Liam HEINIGER a, Norm GOOD b and Sankalp KHANNA b

a University of Queensland, Brisbane, Australia
b The CSIRO Australian e-Health Research Centre, Brisbane, Australia
Patient Flow @ CSIRO AEHRC
Enabling hospitals to better manage their resources & hence reduce waiting times

- Better bed demand prediction
- ED Length of stay performance
- Disease surveillance
- Linking ambulance, ED and admissions data
- Patient flow visualisation
- Patient flow and hospital occupancy
- Bed configuration
- Adverse event analysis
- Early discharge strategies
- Readmission prediction (frequent-flyers)

www.csiro.au/patientflow
Project Background

• National Emergency Access Target (NEAT): The percentage of patients who present to the Emergency Department and are waiting for more than four hours

• Hospital Standardised Mortality Rate (HSMR): ratio of actual number of deaths to expected number of deaths

• Looking at the relationship between NEAT and HSMR

• Our focus here:
  • Predicting the probability of death for patients using Statistical Modelling
UNDERSTANDING THE PROBLEM

Emergency Department

Hospital

4 Hour Breach
Problem Complexity

• Problem at hand - Building statistical models of HSMR
  • Model and predict probability of in-hospital mortality for a given patient
  • HSMR = [Actual number deaths] / [Expected number deaths]
  • Data: Emergency Department (ED) and Inpatient Admission records from several Australian Hospitals over several years
    • In excess of 20 million ED Records.
    • In excess of 20 million Inpatient Records.
    • Large sets of multicollinear variables and potential complex interactions
    • Categorical variables consisting of hundreds of sparsely populated levels.

• Initial Approach
  • Apply a Binomial Generalised Linear Model
  • Intel E5-2630 CPU machine with 2x2.6GHz processors and 128GB of RAM
  • Infeasible solution requiring an unreasonable amount of time and processing power to compute variable estimates.
The Solution

Regularisation – address multicollinearity, reduce number of predictors
• Statistical technique for tuning or selecting the preferred level of model complexity so that models are better at predicting (generalizing).
• Employed Elastic net regularisation
  • Hybrid of 2 popular techniques
  • Increases grouping
  • Reduces coefficients to zero
  • Works well with highly correlated predictors

Variable Aggregation
• Reduce number of categories
• Reduce sparsity
The Solution

Step 1 : Pre-aggregation
• Diseases where all patients died – Highest Risk group
• Diseases where all patients survived – Lowest Risk group

Step 2 : Regularisation
• Parameter estimates for remaining levels determined
  • Using binomial generalised linear modelling
  • Using Elastic Net modelling (cut-off is 1 standard deviation from the minimum error)

Step 3 : Aggregation
• Parameter estimates aggregated into natural bins using the Jenks natural breaks algorithm
Results from Step 2 – GLM Model

Without Pre-aggregation

AUC = 0.75

After Pre-aggregation
Results from Step 2 – Elastic Net Model

AUC = 0.65

75% less time
Results from Step 3

- Parameter estimates aggregated into natural bins using the Jenks natural breaks algorithm
- Calculated parameter estimates placed back into a larger model with the other variables and second order interactions

GLM Model
- AUC = 0.85

Elastic Net
- More ICD-10 codes placed in the “all survival” level.
- AUC = 0.85

The method chosen for aggregating variables is less significant than the act of aggregation itself
Summary

• Complexity often confounds health data modelling
  • Multicollinearity
  • High number of levels in categorical variables

• Conventional models often fail due to such issues

• Techniques like Elastic Net regularisation and variable aggregation can provide efficient mechanisms
Thank you

For more information, please contact:

Norm Good
Senior Experimental Scientist
t  +61 7 3253 3640
e  Norm.Good@csiro.au
w  www.aehrc.com

Sankalp Khanna
Research Scientist
t  +61 7 3253 3629
e  Sankalp.Khanna@csiro.au
w  www.aehrc.com