Neural Networks for Automobile Insurance Pricing

Ai Cheo Yeo
Monash University, Australia

INTRODUCTION

In highly competitive industries, customer retention has received much attention. Customer retention is an important issue, as loyal customers tend to produce greater cash flow and profits, are less sensitive to price, bring along new customers and do not require any acquisition or start-up costs.

BACKGROUND

Various techniques have been used to analyse customer retention. Eiben, Koudijs and Slisser (1998) applied genetic programming, rough set analysis, Chi-square Automatic Interaction Detection (CHAID) and logistic regression analysis to the problem of customer retention modelling, using a database of a financial company. Models created by these techniques were used to gain insights into factors influencing customer behaviour and to make predictions on customers ending their relationship with the company. Kowalczyk and Slisser (1997) used rough sets to identify key factors that influence customer retention of a mutual fund investment company. Ng, Lui and Kwah (1998) integrated various techniques such as decision-tree induction, deviation analysis and multiple concept-level association rules to form an intuitive approach to gauging customers' loyalty and predicting their likelihood of defection.

Mozer and his co-researchers (2000) explored techniques from statistical machine learning to predict churn and based on these predictions to determine what incentives should be offered to subscribers of wireless telecommunications to improve retention and maximise profitability of the carrier. The techniques included logit regression, decision trees, neural networks and boosting. Besides Mozer and his co-researchers, others have also applied neural networks to churn prediction problems. Behara and Lemmink (1994) used the neural network approach to evaluate the impact of quality improvements on a customer's decision to remain loyal to an auto-manufacturer's dealership. Wray and Bejou (1994) examined the factors that seem to be important in explaining customer loyalty. They found that neural networks have a better predictive power than the conventional analytic techniques such as multiple regression. Smith, Willis and Brooks (2000) also found that neural networks provided the best results for classifying insurance policy holders as likely to renew or terminate their policies compared to regression and decision tree modelling.

PREDICTING RETENTION RATES

We have also used neural networks to learn to distinguish insurance policy holders who are likely to terminate their policies from those who are likely to renew in order to predict the retention rate prior to price sensitivity analysis. Policy holders of an Australian motor insurance company are classified into 30 risk groups based on their demographic and policy information using k-means clustering (Yeo, Smith, Willis & Brooks, 2001, 2003). Neural networks are then used to model the effect of premium price change on whether a policy holder will renew or terminate his or her policy. A multilayered feedforward neural network was constructed for each of the clusters with 25 inputs and 1 output (whether the policy holder renews or terminates the contract).

Several experiments were carried out on a few clusters to determine the most appropriate number of hidden neurons and the activation function. Twenty hidden neurons and the hyperbolic tangent activation function were used for the neural networks for all the clusters. A uniform approach is preferred to enable the straight-forward application of the methodology to all clusters, without the need for extensive experimentation by the company in the future. Input variables that were skewed were log transformed.

Some of the issues we encountered in using neural networks to determine the effect of premium price change on whether a policy holder will renew or terminate his or her policy were:

- Determining the threshold for classifying policy holders into those who terminate and those who renew
- Generating more homogenous models
- Clusters that had too few policy holders to train the neural networks

Determining Threshold

The neural network produces output between zero and one, which is the probability that that a policy holder will terminate his or her policy. Figure 1 shows the probability of termination of Cluster 11. A threshold value is used to decide how to categorise the output data. For example a threshold of 0.5 means that if the probability of termination is more than 0.5, then the policy will be classified as terminated. Usually the decision threshold is chosen to maximise the classification accuracy. However, in our case we are more concerned with achieving a predicted termination rate that is equal to the actual termination rate. This is because we are more concerned with the performance of the portfolio (balancing market share with profitability) rather than whether an individual will renew or terminate his or her policy. The actual termination rate for cluster 11 is 14.7%. To obtain a predicted termination rate of 14.7%, the threshold was set at 0.204 (see Figure 1). The confusion matrix for a threshold of 0.204 is shown in Table 1. The overall classification accuracy is 85.3%.

Generating more homogeneous models

The confusion matrix provides the prediction accuracy of the whole cluster. It does not tell us how a given percentage change in premium will impact termination rates. To determine how well the neural networks were able to predict termination rates for varying amounts of premium changes, the clusters were then divided into various bands of premium as follows: decrease in premiums of less than 22.5%, premium decrease between 17.5% and 22.5%, premium decrease between 12.5% and 17.5% and so on. The predicted termination rates were then compared to the actual termination rates. For all the clusters the prediction accuracy of the neural networks starts to deteriorate when premium increases are between 10% and 20%. Figure 2 shows the actual and predicted termination rates for one of the clusters (Cluster 24).

In order to improve the prediction accuracy, the cluster was then split at the point when prediction accuracy starts to deteriorate. This is to isolate those policy holders with a significant increase in premium. It is believed that these policy holders behave differently due to a greater number of these policy holders who have upgraded their vehicles. Two separate neural networks were trained for each cluster. The prediction accuracy improved significantly with two neural networks as can be seen from Figure 3. The average absolute deviation decreased from 10.3% to 2.4%.

Combining Small Clusters

Some of the smaller clusters had too few policy holders to train the neural networks. We grouped the small clusters that had fewer than 7,000 policies. The criterion for grouping was similarity in risk. Risk in turn is measured by the amount of claims. Therefore the clusters were grouped according to similarity in claim cost. The maximum difference in average claim cost per policy was no more than $50. For the combined clusters, prediction ability is also improved by having two neural networks instead of one for each cluster.

PRICE SENSITIVITY ANALYSIS

Having trained neural networks for all the clusters, sensitivity analysis was then performed on the neural networks to determine the effect of premium changes on termination rates for each cluster. There are several ways of performing the sensitivity analysis:

One approach is based on systematic variation of variables (SVV). To determine the impact that a particular input variable has on the output, we need to hold all the other inputs to some fixed value and vary only the input of interest while we monitor the change in outputs (Anderson, Aberg & Jacobsson, 2000; Bigus, 1996).

A more automated approach is to keep track of the error terms computed during the backpropagation step. By computing the error all the way back to the input layer, we have

Figure 1. Determining the threshold value of the neural network output