

Modelling the Effect of Premium Changes on Motor Insurance Customer Retention Rates Using Neural Networks

Ai Cheo Yeo¹, Kate A. Smith¹, Robert J. Willis¹, and Malcolm Brooks²

¹School of Business Systems, Monash University, Clayton, Victoria 3800, Australia
{ai.cheo.yeo, kate.smith, rob.willis@infotech.monash.edu.au}

²Australian Associated Motor Insurers Limited
mbrooks@aami.com.au

Abstract. This paper describes a neural network modelling approach to premium price sensitivity of insurance policy holders. Clustering is used to classify policy holders into homogeneous risk groups. Within each cluster a neural network is then used to predict retention rates given demographic and policy information, including the premium change from one year to the next. It is shown that the prediction results are significantly improved by further dividing each cluster according to premium change. This work is part of a larger data mining framework proposed to determine optimal premium prices in a data-driven manner.

1. Introduction

Like many competitive industries, the insurance industry is driven by two main concerns: the need to return a profit to their shareholders and investors, and the need to achieve market growth and retain a certain level of market share. These two goals are seen as imperatives to success, but are often conflicting. Premium prices play a critical role in enabling insurance companies to find a balance between these two goals. The challenge is to set premium prices so that expected claims are covered and a certain level of profitability is achieved, yet not to set premium prices so high that market share is jeopardized as consumers exercise their rights to choose their insurers.

Insurance companies have traditionally determined premium prices by assigning policy holders to pre-defined groups and observing the average behaviour of each group. The groups are formed based on industry experience about the perceived risk of different demographic groups of policy holders. With the advent of data warehouses and data mining however comes an opportunity to consider a different approach to premium pricing: one based on data-driven methods. By using data mining techniques, the aim is to determine optimal premiums that more closely reflect the genuine risk of individual policy holders as indicated by behaviours recorded in the data warehouse.

In previous work we have proposed a data mining framework for tackling this problem (Smith *et al.*, 2000) (Yeo *et al.*, 2001). This framework comprises

components for determining risk groups, determining the sensitivity of policy holders to premium change, and combining this information to determine optimal premium prices that appropriately balance profitability and market share. Recently we have presented the results of the first component where clustering techniques are used to define risk groups (Yeo *et al.*, 2001). The purpose of this paper is to investigate the second component of the data mining framework: modelling the effect of premium price changes on the customer retention rate.

In Section 2 we review the data mining framework employed in this study. A case study approach utilising a database of over 330,000 policy holders is used to evaluate the effectiveness of various techniques within this framework. Section 3 summarises the previously published results for risk group classification. The use of neural networks for modelling retention rates under various premium changes is then discussed in Section 4. A strategy for improving the retention rate prediction by dividing the data into more homogeneous groups and using separate neural network models for each group is presented, and the results are compared to a single neural network model. Computational results of prediction error rates are presented in Section 5 for all risk classification groups. Conclusions are drawn in Section 6, where future research areas are identified.

2. A Data Mining Framework

Figure 1 presents a data mining framework for determining the appropriate pricing of policies based upon the interaction of growth, claims and profitability. The framework consists of four main components: identifying risk classifications, predicting claim costs, determining retention rates, and combining this information to arrive at optimal premiums. Firstly, the estimated risk of policy holders must be calculated, and used to determine optimal premium values. The total premiums charged must be sufficient to cover all claims made against the policies, and return a desired level of profit. The levels of predicted claims can also be used to forecast profits, when coupled with premium information. However premiums cannot be set at too high a level as customers may terminate their policies, affecting market share. Sales forecasting is determined by marketing information as well as models that predict customer retention or “churn” rates. When integrated, this information provides a methodology for achieving the two goals of market growth and profitability.

For optimal premiums to be set, the insurance company thus needs to determine estimated claim costs and the effect of changes in premiums on retention rates. The estimation of claim cost requires an accurate assessment of risk, discussed in the next section.

3. Risk Classification

Insurance companies group policy holders into various risk groups based on factors which are considered predictors of claims. For example younger drivers are considered to be a higher risk and so they are charged a higher premium. In designing

the risk classification structure, insurance companies attempt to ensure maximum homogeneity within each risk group and maximum heterogeneity between the risk groups. This can be achieved through clustering. In previous work (Yeo *et al.*, 2001) we have shown that the data-driven k-means clustering approach to risk classification can yield better quality predictions of expected claim costs compared to a previously published heuristic approach.

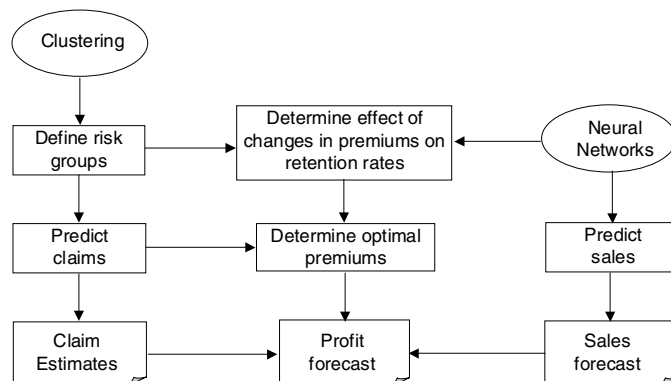


Fig. 1. A framework for data mining within the insurance industry to determine optimal premium prices. It includes components for ascertaining the perceived risk of policy holders, their sensitivity to premium changes, and a mechanism for combining this information to arrive at premiums that try to balance profitability and market share.

The data for this study were supplied by an Australian motor insurance company. Two data sets (training set and test set), each consisting of 12-months of comprehensive motor insurance policies and claim information were extracted. The training set consisted of 146,326 policies with due dates from 1 January to 31 December 1998 while the test set consisted of 186,658 policies with due dates from 1 July 1998 to 30 June 1999. Thirteen inputs were used to cluster policy holders.

The k means clustering model was used to generate a total of 30 risk categories. Figure 2 is a graphical representation of the claim frequency and average claim cost of the 30 clusters.

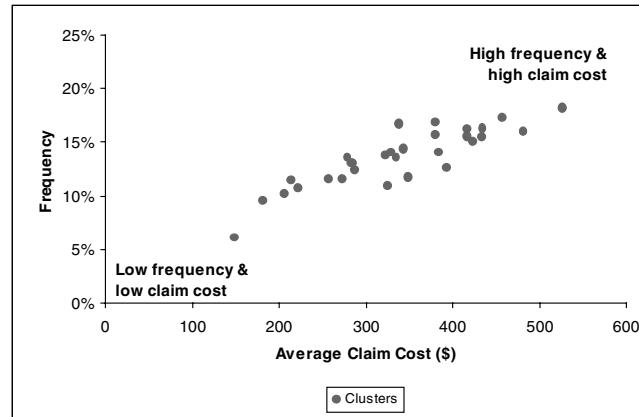


Fig. 2. Claim frequency and average claim cost of the 30 risk groups (clusters)

In the insurance industry, risk is measured by frequency or the probability of a claim and the amount of claims. The higher the frequency, the higher the risk. The higher the amount of claims, the higher the risk. The k-means clustering model was able to find clusters which have significantly different claim frequency and claim cost, without being provided with any claim information as input variables. In other words, clustering is able to distinguish between low and high risk groups.

Besides the k means clustering we also experimented with a previously published heuristic model (Samson *et al.*, 1987) and fuzzy c-means model, but the k-means clustering model produced the best results on the test set with a weighted absolute deviation of 8.3% compared to the 15.6% for the fuzzy c-means model and 13.3% for the heuristic model (Yeo *et al.*, 2001).

Now that the policy holders have been classified into 30 risk groups based on their demographic and policy information, we can examine the pricing sensitivity within each cluster. Neural networks are used in the following section to model the effect of premium price change on retention rates.

4. Modelling Retention Rates

Neural networks have been used in this paper to learn to distinguish policy holders who are likely to terminate their policies from those who are likely to renew. They are an ideal tool for solving this problem due to their proven ability to learn to distinguish between classes, and to generalise their learning to unseen data. Prediction of termination rates or “churn” prediction is a significant area of research, particularly in the telecommunications and insurance industries. Several researchers have successfully applied neural networks to churn prediction problems, and to better understand the factors affecting a customer’s decision to churn (Smith *et al.*, 2000) (Mozer *et al.*, 2000) (Madden *et al.*, 1999) (Behara *et al.*, 1994).

4.1 Data

The training set consisted of 146,326 policies with due dates from 1 January to 31 December 1998 while the test set consisted of 186,658 policies with due dates from 1 July 1998 to 30 June 1999. The period of overlap was to enable comparison of exposure and retention rates over a one-year period. Forty percent of the policies in the test set were new policies. The training set was used to train the neural networks and while the test set was used to evaluate the results.

In addition to the thirteen variables used for risk classification, premium and sum insured information were used as inputs to the neural networks.

4.2 Neural Network Models

A multilayered feedforward neural network was constructed for each of the clusters with 23 inputs, 20 hidden neurons and 1 output neuron. The 23 inputs were:

1. policy holder's age,
2. policy holder's gender,
3. area in which the vehicle was garaged,
4. rating of policy holder,
5. years on current rating,
6. years on rating one,
7. number of years policy held,
8. category of vehicle,
9. sum insured,
10. total excess,
11. vehicle use,
12. vehicle age,
13. whether or not the vehicle is under finance.
14. old premium,
15. new premium,
16. old sum insured,
17. new sum insured,
18. change in premium,
19. change in sum insured,
20. percentage change in premium,
21. percentage change in sum insured,
22. ratio of old premium to old sum insured,
23. ratio of new premium to new sum insured.

The hyperbolic tangent activation function was used. Input variables which were skewed were log transformed. The neural network produces output between zero and one, which is the probability that a policy holder will terminate his policy. Figure 3 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. In our case, we have set a threshold so that the predicted termination rate is equivalent to the actual termination rate for the whole cluster. For cluster 11, the

termination rate is 14.7% which means a threshold of 0.204 is needed to produce the same predicted termination rate, as shown in Figure 3.

4.3 Prediction Accuracy

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% etc. 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% to 20%. Figure 4 shows the actual and predicted termination rates for one of the clusters (Cluster 24).

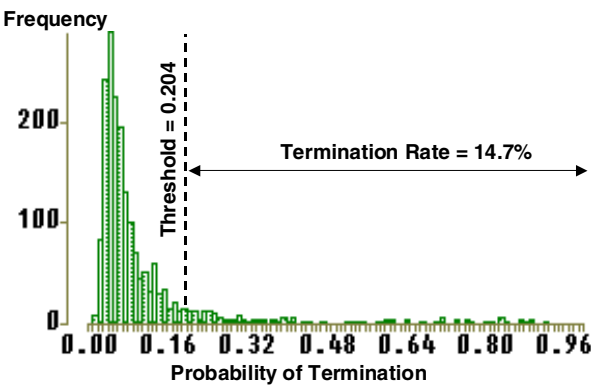


Fig. 3. Determining the threshold value of the neural network output

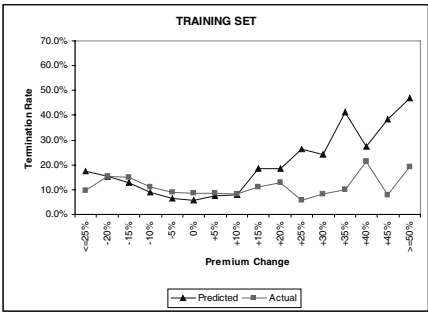


Fig. 4. Prediction accuracy for one neural network model of cluster 24

4.4 Generating More Homogeneous Models

In order to improve the prediction accuracy, the cluster was then split at the point when prediction accuracy starts to deteriorate. Two separate neural networks were trained for each cluster. The prediction accuracy improved significantly with two neural networks as can be see from Figure 5. The average absolute deviation decreased from 10.3% to 2.4%.

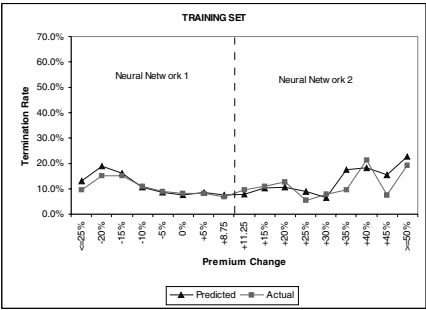


Fig. 5. Prediction accuracy for two networks model of cluster 24 (training set)

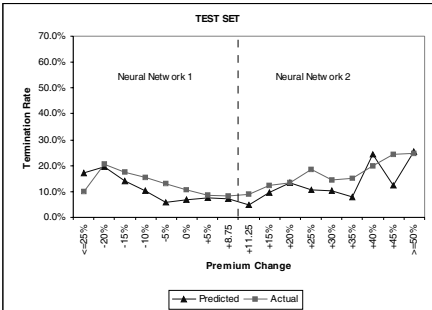


Fig. 6. Prediction accuracy for two networks model of cluster 24 (test set)

The neural network was then applied to the test set. The neural network performed reasonably well on the test set with an average absolute deviation of 4.3% (Figure 6).

4.5 Combining Small Clusters

There were some small clusters which have too few policy holders to train the neural networks. We grouped the small clusters which had fewer than 7,000 policies. Since the objective is to ultimately determine the optimal premium which reflect the risk of the policy holders, the criteria for grouping has to be 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. Table 1 shows the grouping of the small clusters.

For the combined clusters, prediction ability is also improved by having two neural networks instead of one for each cluster as can be seen from Figure 7 and Figure 8. The average absolute deviation decreased from 10.3% to 3.5%. The test set has an absolute deviation of 4.2% (Figure 9).

5. Results

Table 2 presents a summary of the results for all clusters. It shows clearly that the average absolute deviation between the actual and predicted termination rates is significantly reduced by employing two neural networks per cluster rather than a

single neural network. It appears that a single neural network is unable to simultaneously learn the characteristics of policy holders and their behaviours under different premium changes. This is perhaps due to the fact that many of the large premium increases are due to an upgrade of vehicle. Since these policy holders may well expect an increase in premium when their vehicle is upgraded they may have different sensitivities to premium change compared to the rest of the cluster. Attempting to isolate these policy holders and modelling their behaviours results in a better prediction ability.

Table 1. Grouping of Small Clusters

| Cluster | No of Policies | Average Claim Cost | Difference | |
|---------|----------------|--------------------|------------|----|
| 13 | 2,726 | 344 | } | 16 |
| 14 | 2,714 | 343 | | |
| 28 | 6,441 | 328 | | |
| 12 | 1,422 | 285 | } | 7 |
| 2 | 3,988 | 280 | | |
| 30 | 2,366 | 278 | | |
| 5 | 5,606 | 270 | } | 14 |
| 26 | 1,460 | 262 | | |
| 23 | 3,374 | 256 | | |
| 3 | 1,595 | 249 | } | 15 |
| 10 | 1,610 | 248 | | |
| 8 | 2,132 | 247 | | |
| 7 | 1,601 | 235 | | |
| 18 | 1,446 | 234 | | |
| 1 | 4,101 | 231 | } | 48 |
| 15 | 1,621 | 217 | | |
| 16 | 1,505 | 194 | | |
| 6 | 2,346 | 183 | | |
| 21 | 3,566 | 138 | } | 40 |
| 25 | 1,445 | 125 | | |
| 22 | 1,401 | 116 | | |
| 17 | 1,411 | 98 | | |

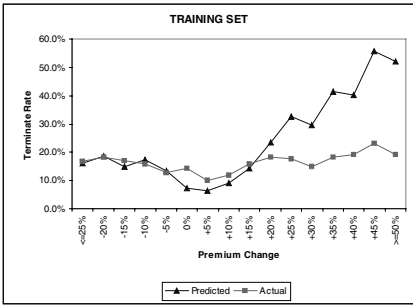


Fig. 7. Prediction accuracy for one network model of combined clusters 5, 23 and 26

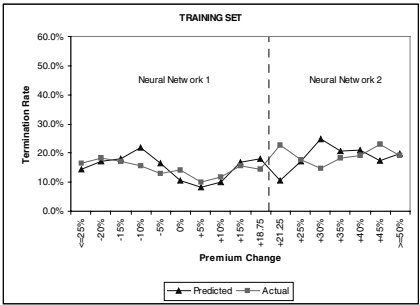


Fig. 8. Prediction accuracy for two networks model of combined clusters 5, 23 and 26 (training set)

6. Conclusions and Future Research

This paper has examined the use of neural networks for modelling customer retention rates within homogeneous groups. The work is part of a data mining framework for determining optimal premium prices. Clustering is used to arrive at homogeneous groups of policy holders based on demographic information. This information is then supplemented with premium details, and a neural network is used to model termination rates given premium changes. We have shown that significant improvements in prediction accuracy can be obtained by further dividing each cluster to isolate those policy holders with a significant increase in premium. It is believed that these policy holders behave differently due to the greater number of these policy holders who have upgraded their vehicles.

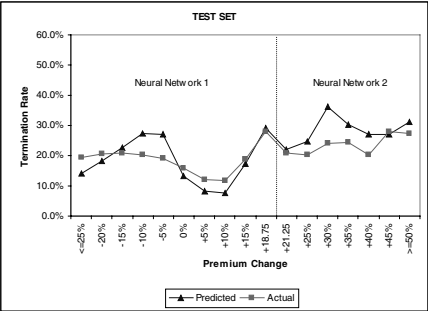


Fig. 9. Prediction accuracy for two networks model of combined clusters 5, 23 and 26 (test set)

Table 2. Summary of Results

| Cluster | ONE NETWORK | | TWO NETWORKS | | | | |
|-----------------|----------------------|----------------------------------|--------------|----------------------------------|-----------|--------|----------------------------------|
| | TRAINING SET | | Split at | TRAINING SET | TEST SET | | |
| | Actual/ Predicted | Average Absolute Deviation | | Average Absolute Deviation | Predicted | Actual | Average Absolute Deviation |
| | | | | | | | |
| 4 | 11.2% | 5.5% | 20% | 2.7% | 9.9% | 12.6% | 2.9% |
| 20 | 6.8% | 10.8% | 15% | 3.7% | 5.0% | 6.9% | 3.4% |
| 11 | 14.7% | 8.3% | 10% | 5.0% | 17.1% | 17.1% | 7.4% |
| 29 | 9.5% | 8.8% | 15% | 3.8% | 8.5% | 11.3% | 5.9% |
| 9 | 8.8% | 7.2% | 20% | 1.9% | 8.6% | 10.0% | 4.5% |
| 24 | 9.5% | 10.3% | 10% | 2.4% | 8.2% | 10.9% | 4.3% |
| 27 | 11.6% | 12.4% | 15% | 3.1% | 10.8% | 12.6% | 4.2% |
| 19 | 7.5% | 6.3% | 15% | 2.2% | 8.1% | 7.6% | 3.6% |
| 13, 14, 28 | 15.1% | 10.9% | 10% | 3.8% | 17.9% | 17.9% | 4.8% |
| 2, 12, 30 | 13.5% | 9.2% | 10% | 2.8% | 13.1% | 15.4% | 4.0% |
| 5, 23, 26 | 14.7% | 10.3% | 20% | 3.5% | 17.7% | 17.3% | 4.2% |
| 3, 7, 8, 10, 18 | 11.7% | 6.7% | 20% | 3.1% | 12.3% | 14.4% | 4.4% |
| 1, 6, 15, 16 | 10.9% | 5.2% | 20% | 2.2% | 11.1% | 12.9% | 4.0% |
| 17, 21, 22, 25 | 8.7% | 5.4% | 15% | 2.9% | 8.4% | 9.8% | 3.4% |

References

- Behara, R.S., Lemmink, J. (1994): Modeling the impact of service quality on customer loyalty and retention: a neural network approach. 1994 Proceedings Decision Sciences Institute. 3, 1883-5
- Madden, G., Savage, S., Coble-Neal, G. (1999): Subscriber churn in the Australian ISP market. *Information Economics & Policy* 11(2), 195-207
- Mozer, M.C., Wolniewicz, R., Grimes, D.B., Johnson, E., Kaushansky, H. (2000): Predicting subscriber dissatisfaction and improving retention in the wireless telecommunication. *IEEE Transactions on Neural Networks* 11(3), 690-6
- Samson, D., Thomas, H. (1987): Linear models as aids in insurance decision making: the estimation of automobile insurance claims. *Journal of Business Research* 15, 247-256
- Smith, K.A., Willis, R.J., Brooks, M. (2000): An analysis of customer retention and insurance claim patterns using data mining: a case study. *Journal of the Operational Research Society* 51, 532-541
- Yeo, A.C., Smith, K.A., Brooks, M. (2001): A comparison of soft computing and traditional approaches for risk classification and claim cost prediction in the automobile insurance industry. In: Reznik, L., Kreinovich, V. (Eds.): *Soft computing in measurement and information acquisition*. Springer-Verlag, Heidelberg, accepted for publication.