Knowledge Discovery on Customer Churn Prediction

LI-SHANG YANG [1], CHAOCHANG CHIU [2]

[1] Department of Business Administration, St .John's University, Tamsui, Taiwan, ROC

[2] Department of Information Management, Yuan Z University, Chungli, Taiwan, ROC

Abstract: - As **customer churn** has become a critical business issue for most mobile telecommunications providers, predictive modeling based on knowledge discovery of data mining is an approach being used to facilitate customer retention more effectively and proactively. This paper reports a preliminary study of applying data mining to solve a real-world customer attrition problem. It uses customer information to make predictions about the likelihood of churn. The target rate is very small, around 0.5% monthly churn rate. Nevertheless, by scrutinizing data to extract a large number of independent variables and by using a large training data set, the predictive model delivers excellent performance in field test. Issues in the practical application of data mining are also discussed.

Key-Words: - Customer Retention, Churn Prediction, Data Mining, Decision tree, Field Test, Logistic Regression, Telecommunication.

1 Introduction

It is acknowledged that one of the weakest areas of organizational performance by quality award entrants is that of 'data and information'. In particular, information and knowledge about customers is central to maintaining a quality service focus.

Knowledge derived from social capital in the form of information about customers, and the development of organizational capital through databases and proprietary processes for the treatment of this data (e.g. data mining models), is an important part of building a competitive organization.

This paper describes the first efforts of a major Taiwanese telecommunications provider in this area. The company knew that the market situation in which they were operating was unheralded anywhere in their industry [3]. From a base rate of only 7%, the penetration rate (mobile users per 100 people) had escalated to 106.15% within 5 years.

2 Customer Churn in Mobile Service

This study focuses on post-paid customer voluntary churn prediction for a major Taiwan mobile service operator. In the telecom service industry, churn means customer attrition [8]. Churn can be of several types:

- involuntary churn: This occurs when subscribers fail to pay for service and as a result the provider terminates service. Termination of service due to theft or fraudulent usage is also classified as involuntary churn;
- Unavoidable churn: This occurs when customers die, move or are otherwise permanently removed from the market place;
- Voluntary churn: Service termination on the part of the customer when leaving one operator and possibly for another because of better value

In reality, it is very unlikely that a service provider could differentiate unavoidable and voluntary churn and predict them separately. We combine the two. As for involuntary churn, most Taiwan mobile service providers allow at least several months' "grace period" before cutting off delinquent customers. Thus, the business value of predicting involuntary churn is limited.

3 Related Research on Customer Churn

A range of researchers have used experimental statistical approaches and data-mining techniques for customer churn prediction modelling in industries such as mobile telephony, banking, insurance and internet service provision[1,2,4,6,7,9,10,12]. A review of their research, the churn rate ranged from 5% to 20% and the number of records from 592 to 5 million. In only two cases is there reporting of the churn mix in the training set. In one case this was placed at 50%[7] and in the other at 6.2%[9]. In three

cases 'Lift' is used as the method to evaluate performance [7,9,10], for others, the performance measurement is unavailable. However, none of them report the Lift value of their final model, and hence it is difficult to compare the performance of their models.

The modeling techniques of these research included decision trees, neural networks, regressions, and binomial probability. All researchers confirm the contribution of corresponding techniques to their modeling results.

None of studies discusses the use of a field test of model performance in an actual organizational field environment. Typically training and testing of models is done with historical data within the same time span. A field test traces the outcome of a predicted value in the real word at a time span different from training and testing set. A field test helps to examine the validity and reliability of the model.

Field testing is an important process in verifying the ultimate practical application of this type of approach. Model accuracy is important, and the use of unreliable model may cause significant monetary loss in business. We enhanced our study in a novel way by conducting the field testing, and this was not found to be reported in the past research.

4 Data Mining Role in Retention Process Model

Customer Retention is a "business process," with the purpose to win the loyalty of customers. Fig1 shows a conceptual framework of customer retention process, KCRPM, developed by the author [8].

Knowledge discovery is a sub-process of Knowledge Management Process (Fig 1) that provides:

- Insights as to what kind of customers are leaving and possible causes. This is a "profiling" process.
- Lists of customers who are about to leave. This is a "prediction" process.

Relationship-Marketing Planning Process enables the development of retention strategic plans; predefines specific targets, treatment plans, and campaign types, and convert plans into actions based on knowledge gained in Knowledge Management process.. Relationship-Marketing Implementation Process takes action on sales and service plans delivered. Through channels, results information of execution will be collected and feedback for performance evaluation. The performance measurements provide information for learning and refinement. The Knowledge Management Process refines the knowledge and makes the information available for sharing. This in turn enables process improvement and new knowledge formation. This forms a process cycle of continuous learning and improvement.

This study is, to an extent, a validation of the conceptual framework presented in Figure1. The focus is on the Churn prediction; data mining are applied to build prediction models so that specific retention targets for retention strategic plans in Relationship-Marketing Planning Process can be defined.

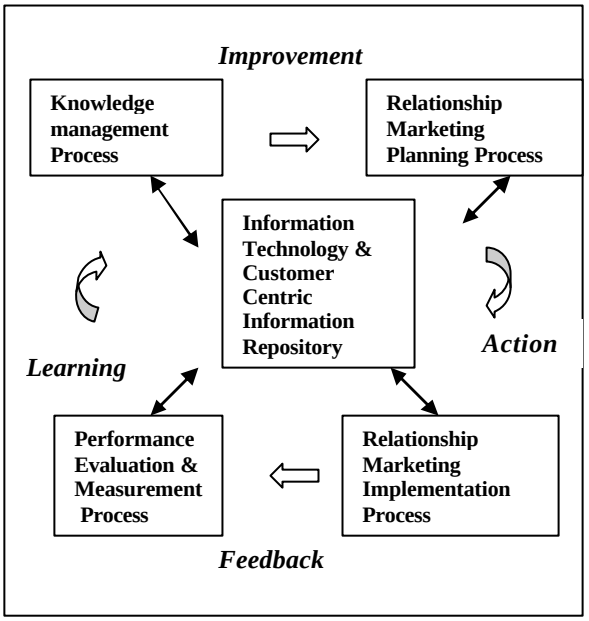

Fig 1: Knowledge-based Customer Retention Process Model [5]

5 Predictive Modeling Process

5.1 Experiment Architecture Subsection

Fig 2 shows the experiment architecture in this study which was partitioned into modules interacting with the data warehouse.

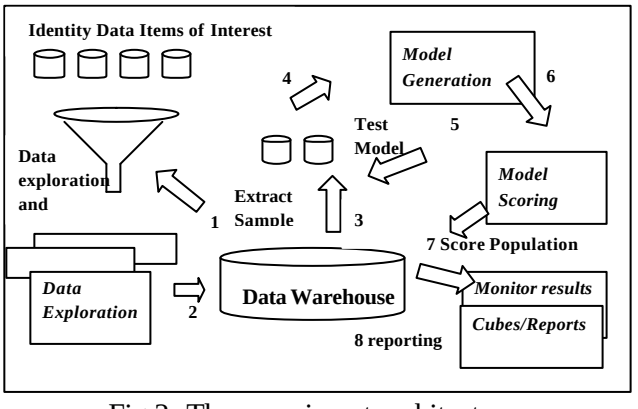

Fig 2: The experiment architecture

Step 1, customer data in the data warehouse was extracted and explored to identify data items of interest. Step 2 involved the designing of programming scripts to extract and transform data into the desired form after the variables of interest were selected. These data were restored in the data warehouse. In Step 3 data was sampled extracted for further analysis, churn prediction and value modeling. Steps 4 and 5 comprised the recursive interaction with the data warehouse until models were fine tuned. In Steps 6 and 7, SQL scripts were designed to automate and transfer the discover rules of model into scores. The scoring program was applied to entire customer population in the data warehouse to produce churn risk and value scores for each customer on a monthly basis. In Step 8, a list of 5000 high churn propensity customers accompanying with their value score was generated. Other knowledge (e.g. customer segments) were presented in the form of OLAP cubes; and monitor reports presented lift information which helped to monitor for model degradation beyond the user-defined threshold.

The following success criteria were set for the experiment:

- Model performance - Hit rate and Lift was used as indicators. Essentially, a field test was used to measure the stability of models over time.
- Model Interpretability - Are the rules generated by the model understood and agreed with by marketing users?
- Model efficiency - The time taken to generate a model.
- Model Maintainability - The ease of development or refinement of the model.

Critical dimensions of the study included: defining churn as voluntary churn, the prediction window was 1 month, "Qualified Subscribers," were used (i.e. those with a post-paid contract with 4 months or longer tenure and with selected bill plans), and the name list predicted size was kept to 5000 to accommodate outbound telemarketing staff constraints

.

5.2 Variables Selection

The data warehouse integrates data about products and services, customer billing, customer care summaries, contract, and promotion data, including rated CDR (Call Detail Record). the "on-line" data warehouse consists of about 6 million mobile service subscribers.

Over 170 variables were identified based on customer demographics, channel, and contractual attributes; measurements, such as usage, invoices, payments; and events, such as changes, late payments, loss of SIM cards, and customer care contacts.

Each variable was visually explored in a search for obvious patterns. If a pattern was detected was tested with a chi-square test. Any variable whose univariate test had a p-value <0.25 was considered as a candidate for inclusion in the prediction model. The level of 0.25 was used instead of the traditional 0.05 to allow more variables to be selected in this stage. In addition to the chi-square test, odds ratio was also used to identify significant categorical explanatory variables.

For continuous variables (such as age), we use entropy to dissect the domain for best separation between churn and in-use customers. Variables were combined to derive new ones that are more powerful in differentiating churn and in-use customers. Eventually the variable list was reduced to 99.

6 Modeling and Pattern Recognition

There are many prediction models in statistics and data mining that can be applied to churn prediction problems. Past research shows that Decision Tree, Neural Network, Logistic Regression are well suited to this type of problem. However, neural network analysis provides internal weights which are distributed throughout the network. As these weights don't provide insight into why the solution is valid; they are not readily understandable by the end user. Neural networking is often cited as a methodology that builds a black box. In comparison, tree models are characterized by "simple" if-then rules that can sometimes be easy to understand. Regression models usually provide a straightforward relationship between the explanatory variables and the prediction. It is important to choose the algorithm considering the aspect of intellectual 'comfort' issue. If model users understand the relationships being used, and if these relationships make sense to them, then they more readily accept the validity of the model. In the current study this interpretability issue led the team to adopt decision tree and logistic regression approaches.

6.1 The Decision Tree

The decision tree is a machine learning algorithm which learns rules from data. Decision trees are a good choice when the data mining task is classification of records, or prediction of outcomes. Decision trees are also a natural choice when the goal is to generate rules that can be easily understood, explained, and translated into SQL of a natural language.

In the current study, because the monthly churn rate was low (about 0.5%); it was difficult to focus a machine learning system on churners. We required a large training data set and biased the churn/in-use mixes in both training and testing data set. Various mixes from 1 to 10 % were tried with training and test data sets which ranged in size from 50,000 to 1,000, 000 records.

A complex tree with many branches and leaves can be very difficult to understand. Different prune severity and minimum branch size were used to simplify the tree and achieve the best hit rate for the 5000 name list.

To ensure the developed model was not 'over-fitting', that is a tendency to merely memorize the patterns in the training set; another validation data set with 0.55 % churn rate was applied.

6.2 Logistic Regression

Logistic regression has become the standard method for regression analysis of dichotomous data in many fields, especially in the health sciences. A logistic regression model is distinguished from a linear regression model because the outcome variable in the logistic regression is binary or dichotomous. Once this difference is accounted for, the methods employed in an analysis using logistic regression follow the same general principles used in linear regression.

Consider a collection of p independent variables which will be denoted by the vector x' = $(x_1, x_2, \ldots, x_p)$. Let the conditional probability that dependent variable is present be denoted be P(Y=1|x) = $\pi(x)$. Then the multiple logistic regression model is given by the equation. :

$$g(x) = g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

In which case $$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

In this study, where x' = attribute variables (i.e. age, tenure, etc.), $\pi(x)$ = the probability of churn. The Stepwise method was applied in variable selection while modeling. Thirty one (31) variables were finally identified with a $p < 0.001$. A 3% churn mix was used with the training set and the model validated with the data set with a churn rate of 0.55 %.

7 Model Performance

7.1 Measurements: Hit Rate and Lift

For the Decision Tree, different churn mix and sample size were used to train the model, the best model was obtained at churn rate 2% and sample size 375,000. It was therefore decided to choose the model with 2% mix, where prune severity = 75, branch size = 15. It is also found that when the sample size was over 375,000, the hit rate was stable and higher than with smaller sets. Obviously, the larger data set size contributed positively to model performance but plateaus at about this size. For a 5000-name list, the Hit Rate was 56%. That is, 56% of the 5000 names identified did actually churn in the prediction window. The lift value for this 5000 name list was over 100.

As to the Logistic Regression, for a 5000-name list, The Hit Rate was 41% and the lift for the 5000-name list was 74, compared with a random selection of 0.55% population churn rate.

The performance of logistic regression was poorer than decision tree. It was decided to use decision tree as the modeling algorithm in our final modeling application.

7.2 Field Testing

Field testing is evaluating the model' s validity in the real world. Table 1 shows the performance result of applying the decision tree model to successive months based on the 5000 name list. No management intervention aimed to improve retention took place during the test period. The table shows that the model was stable over successive months.

Table 1: Performance results of field test

Test Type	Month to Predict	Hit Rate	Churn Rate	Lift
Validation Test	1' st month	56%	0.55%	102
Field Test	2' nd month	47%	0.51%	92
	3' rd month	35%	0.35%	99
	4' th month	46%	0.48%	96

8 The Findings

The study produced a variety of findings. The knowledge discovered about customers has obvious potential to support marketing efforts. This section highlights a range of issues.

In terms of the effectiveness of the data mining model, even with a low churn rate (0.55%) the model provided lift figures of near 100. Under many circumstances, a lift figure of 10 might be regarded as valuable. The study illustrated that data mining is applicable even when small target rates are in existence.

The well-defined scope, metrics, process ownership, and hard work all contributed to the success of this experiment. Active support from and participation of Customer Care, Marketing, and Billing subject experts has significantly reduced the development interval. A well-integrated data warehouse and powerful servers have reduced model generation time to a couple of hours.

It is believed that a good set of metrics to measure the model was selected, and the model performs pretty well based on these metrics.

Data quality is a perennial issue in this type of exercise. It is recognized that the data can never be perfect. Raw CDR data of call quality was not available. Only summarized customer care transactions were used and it would be an interesting study to see how much the hit rate improves with additional data. A 'quality from start' rule was deliberately applied to data at beginning of the exercise to avoid the trap of garbage in, garbage out.

It was learnt that there are no good rules to determine a good mix to train the machine. A lot of time was spent in trial and error processes.

In terms of model automation, the study implies that it is not likely a turn-key application package can fit in data mining processes in the industry. Knowledge discovery of data mining is different from reporting functions. It involves many iterative decision processes involving human intelligence and judgments. There are parts of the broader process that cannot be automated, including choosing a methodology to match a business problem, selecting a data set, quality checking and preparing the data for analysis, choosing among the available options within the analysis process, and interpreting and presenting the results.

Automating measurement of Lift in regular reports is strongly recommended as an efficient way to monitor the model performance in the real world, and saves monitoring efforts in determining when the model needs refinement. The study does not nominate a model performance threshold under which the model needs retune because any threshold is likely to be application dependent and strongly linked to company's business strategies. Field test is strongly recommended to ensure the validity of the model.

The cost of computing technology is a necessary consideration. The large volume data required for good performance of models suggests that a relatively powerful server might reduce the model training interval.

9 Conclusion

This paper describes and discusses a study in the mobile telephone industry. It demonstrates convincingly the benefits of the application of a knowledge discovery process through data mining for informing decision making on customer defection.

Retaining customers is one of the most important issues in customer relationship management (CRM). It is believed that the data mining process undertaken underpins a sound quality service and customer care improvement approach. The model has allowed the company to identify several areas for enhancement and the challenge is now to incorporate the model into marketing strategies and make retention actions on it.

Data mining provides a collection of methodology, techniques, and approaches that help to fully integrate the solutions with existing business knowledge and implementation, and as such is likely to become a highly useful customer retention tool in the coming years.

References:

[1] Wouter Buchinx, Geert Verstraeten, Dirk Van den Poel, Predicting customer loyalty using the internal transactional database, *Expert Systems with Applications*, Vol. 32, No. 1, 2006, pp. 125-134.

[2] Hung-Chang Chiu, Yi-Ching Hsieh, Yu-Chuan Li, Monle Lee, Relationship marketing and consumer switching behavior, *Journal of Business Research* ,Vol.58, 2005, pp.1681-1689.

[3] Directorate General of Telecommunications, *Mobile subscribers statistic report, 2003* [Online, 1 Dec.2003] http://www.dgt.gov.tw.

[4] John Hadden, Ashutosh Tiwari, RaJkumar Roy, Dymitr Ruta, Computer assisted customer churn management: State-of-the-art and future trends, *Computers & Operations Research*, [Online, 16 Sep.2006] http://www.sciencedirect.com/

[5] IDC, IDC: European mobile market saturated -- Operators must focus on customer retention, says IDC, [Online, assessed August, 2002], http://www.idc.com/itforum02.

[6] Madden, G., Savage, S. J. & Coble-Neal, G., Subscriber churn in the Australian ISP market, *Information Economics and Policy*, vol. 10, 1999, pp. 195-207.

[7] Masand, B., Datta, P., Mani, D. R., Li, B. & GTE Laboratories, W., MA, CHAMP: A Prototype for Automated Cellular Churn Prediction, *Data*

Mining and Knowledge Discovery, Vol. 3, 1999, pp. 219-225.

[8] Modisette, L., Milking Wireless Churn for Profit, *Telecommunications*, 1999, pp. 73-74..

[9] Mozer, M. C., Wolniewicz, R., Grimes, D. B., IEEE, Johnson, E. & Kaushansky, H., Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry, *IEEE Transactions on Neural Networks*, vol. 11, no. 3, 2000, pp. 690-696.

[10] Smith, K., Willis, R. & Brooks, M., An analysis of customer retention and insurance claim patterns using data mining: a case study, *Journal of the Operational Research Society*, vol. 51, 2000, pp. 532-541.

[11] Yang, Li-Shang, Research Paper, *Literature-based Critical Review of Customer Churn Behavior for Mobile Phone Industry in Taiwan, January 2003*, International Graduate School of Management, University of South Australia.

[12] Trubik, E. & Smith, M., Developing a model of customer defection in the Australian banking industry, *Managerial Auditing Journal*, vol. 15, no. 5, 2000, pp. 199-208.