CUSTOMER CHURN MODELS:
A COMPARISON OF PROBABILITY AND DATA MINING APPROACHES

Ali Tamaddoni Jahromi, Monash University, Australia
Stanislav Stakhovych, Monash University, Australia
Michael Ewing, Monash University, Australia

INTRODUCTION

As markets become increasingly saturated, astute companies acknowledge that their business strategies should focus on identifying those customers who are likely to churn (Hadden, Tiwari, Roy, & Ruta, 2007). Since net returns on investments for retention strategies are generally higher than for acquisitions, it is generally accepted that companies should concentrate their marketing resources to keep existing customers rather than to attract new ones (Colgate & Danaher, 2000). This calls for models capable of making accurate predictions about consumers' behavior in a future time period. Such models should be able to specify *which* customers in a dataset have a higher probability to churn in a given future time period. Literature on churn modeling reveals that predictive models fall into one of two categories, namely probability modeling and data mining modeling. Although many studies from both of these streams have focused on developing models to predict and identify customer churn, to the best of our knowledge, none of them have compared the performance of these modeling approaches in terms of accuracy in identifying and predicting customer churn.

On this basis, the current study aims to compare the performance of probability and data mining model building approaches. In this regard, the dataset from a company operating in non-contractual setting has been utilized to construct predictive models of customer churn with both approaches. The models have then been compared in terms of their accuracy in identifying churners. The results suggest that although models constructed with probability and data mining model have the same general accuracy, the decision tree model with cost sensitive learning has the upper hand in identifying the true churners.

BACKGROUND

Probability models

Introduced by Ehrenberg (1959), probability models utilize simple probability distributions to model the observed behavior of individual customers and make predictions regarding their future behavior (Fader & Hardie, 2009). The work of Ehrenberg (1959, 1972) was further extended by Schmittlein, Morisson, and Colombo (1987) and Fader, Hardie, and Lee (2005) to develop Gamma-Exponential / Negative Binomial Distribution (Pareto/NBD) and Beta-Geometric / Negative Binomial Distribution (BG/NBD), respectively.

The Pareto/NBD and BG/NBD form the backbone of almost all studies in customer base analysis and are attractive because they: (1) utilize previous transaction behavior to construct the model; (2) predict the individual's future purchase level; and (3) give the probability that a specific customer is active after a specific time. The collection of these features have made these two models the most well-known and recommended stochastic methodologies for recognition of customer churn as well as prediction of future sales in non-contractual settings (Wübben, 2008).

Data mining models

Data mining is the analysis of large data sets to find patterns and models in the data which summarize the data in a way that is more useful and understandable for the data owner (Hand, Mannila, & Smyth, 2001). The tendency towards employing data mining techniques in customer churn prediction stems from the fact that churn is a rare event in a dataset and making an accurate forecast calls for techniques that emphasize predictive ability (Kamakura et al., 2005). Several studies have constructed different data mining models to predict customer churn in various sectors such as telecommunications (Burez & Van den Poel, 2009), finance (Xie, Li, Ngai, & Ying, 2009), retail (Buckinx & Van den Poel, 2005), and Pay TV (Burez & Van den Poel, 2008). Moreover, recent studies in the marketing literature have emphasized the applicability of machine learning models as an alternative for standard approaches in marketing, like logit models (Cui & Curry, 2005; Lemmens & Croux, 2006).
Data mining techniques have also been acknowledged as a remedy for the limitations of probability models in dealing with real world data. Such constraints have their roots in the fact that probability models are established based on the assumptions about the way the data is distributed and these types of assumptions can be restraining and in some cases misleading

(Wübben, 2008). However, none of the existing studies have empirically compared the performance of data mining models in churn prediction against the probability models. This calls for further investigation to directly compare these two approaches in order to employ the best possible modeling approach when dealing with customer churn.

METHODOLOGY AND RESULTS

The data for this study comes from customer transactional records of the online CD retailer *CDNOW*, in a period between January 1997 and June 1998. The original data which has been used by Fader and Hardie (2001) contains sales records of 23,570 customers. However, the data set that has been used for the current study is a 1/10th sample of the original dataset (2,357 customers) available on Bruce Hardie's website (Fader et al., 2005). In the current study, this sample data set has been divided to 1649:708 (70:30), train set: test set.

The data set contains the transactional records of customers during a 78 week period. For the model building purposes this time window has been broken into two equal 'calibration' and 'validation' periods of 39 weeks, respectively (Figure 1). In order to compare the performance of both modeling approaches the 'recency', 'frequency', and 'monetary value' aspects of the customers' transactions should be specified as follows:

1- x: number of transactions observed between the time of the first purchase (t=0) and end of observation period ($t = T$)
2- t_x: time of the last repeated transaction in observation period ($0 \leq t_x \leq T$)
3- T: known as observation period, is the time between the first period and end of observation period which varies across the customers
4- m_x: the average 'monetary value' of each transaction in observation period
5- Churn: a binary variable which indicates whether a customer who has made his/her first purchase in the calibration period is still active in validation period or not. On this basis, the customers who have made at least one purchase in validation period have been considered as being 'alive' (coded as 0) and customers who have no transaction record in the prediction period have been considered to be 'dead' (coded as 1).

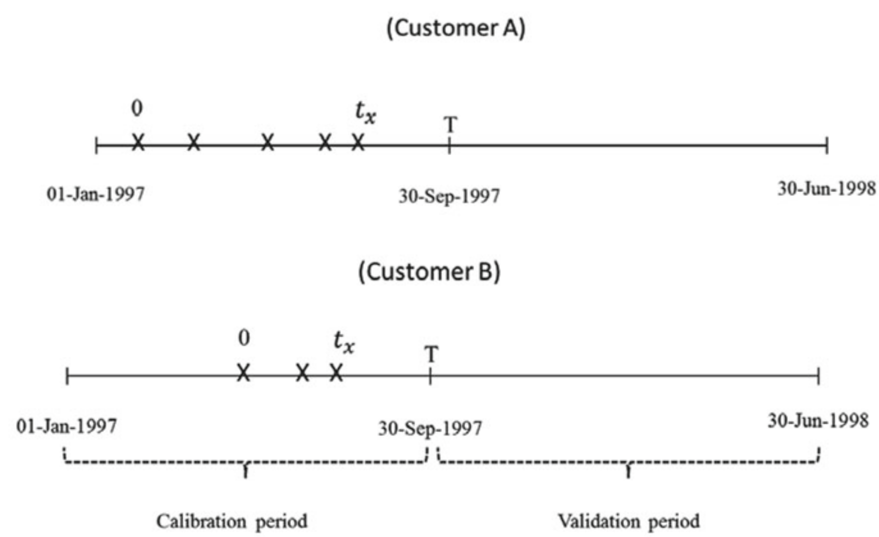

Figure 1 CDNOW modeling timeline for two customers with different observation length

Probability modeling approach

Both BG/NBD and Pareto/NBD models can provide the probability that an individual customer with the purchase history of x, $t_x$and T is 'alive' at the end of calibration period. However, due to the fact that BG/NBD model has problem with calculating P(Alive) for customers with no purchase in calibration period (Fader, Hardie, & Lee, 2008; Hoppe & Wagner, 2007), this study investigates the performance of only Pareto/NBD model.

Setting the four parameters of the model as $r = 0.55$, $\alpha = 10.57$, $s = 0.6$, and $\beta = 11.66$ (Fader et al., 2005), the P(Alive) at time T has been calculated for all customers in test data set using equation 1, provided by Schmittlein et al. (1987).

$$P(\text{Alive}|r,\alpha,s,\beta,x,t_x,T)=\left\{1+\frac{s}{r+x+s}\times\left[\left(\frac{\alpha+T}{\alpha+t_x}\right)^{r+x}\left(\frac{\beta+T}{\alpha+t_x}\right)^{s}{}_2F_1\left(r+s+x,s+1;r+s+x+1;\frac{\alpha-\beta}{\alpha+t_x}\right)-\left(\frac{\beta+T}{\alpha+T}\right)^{s}{}_2F_1\left(r+s+x,s+1;r+s+x+1;\frac{\alpha-\beta}{\alpha+T}\right)\right]\right\}^{-1} \quad Eq.1$$

After calculating the P(Alive) value for all customers in the test set, the cutoff point of 0.5 (Wübben & Wangenheim, 2008) was used to form a binary variable of 'churn', i.e. for $\text{P(Alive)} \geq 0.5$ the customers were considered to be 'alive' (0) and for $\text{P(Alive)} < 0.5$ 'dead' (1). Table 1 reports the error matrix of our model based on the constructed churn variable.

Table 1 Error matrix Pareto/NBD; P(Alive) estimation

Actual	**Predicted**	
	0	**1**
0	108 (15.3%)	99 (14%)
1	68 (9.6%)	433 (61.1%)

Data mining modeling approach

Using x, t_x, T, and m_x as input variables and 'churn' as the target variable, a decision tree model with CART algorithm (Breiman, Friedman, Olshen, & Stone, 1984) was constructed using CDNOW data. With this aim, 70% of the customers in the data set have been allocated to the training set and used to construct the training model. Once the training model was constructed, its performance was tested, using data from the remaining 30% of customers. The decision tree analysis has been chosen over other binary classifiers due to its interpretability and understandability for business people (Ngai, Xiu, & Chau, 2009) along with its transparency as indicated by Olafsson, Li, and Wu (2008).

Table 2 depicts the error matrix of the constructed model tested on our test set of 708 customers:

Table 2 Error matrix decision tree

Actual	**Predicted**	
	0	**1**
0	104 (14.7%)	103 (14.5%)
1	72 (10.2%)	429 (60.6%)

It was shown in the literature that the cost of acquiring a new customer are often up to five times greater than the cost of retaining an existing customer (Kotler & Keller, 2006; Rosenberg & Czepiel, 1984). Therefore, the failure in identifying churners (False Negative error) may cost five times more for the company compared to sending incentives to customers who are not churners (False Positive error). The effect of cost-sensitive learning (Burez & Van den Poel, 2009) on the prediction accuracy of the classifier has been tested and Table 3 illustrates the error matrix for ratio 1:5 (FP: FN).

Table 3 Error matrix decision tree; cost sensitive learning

Actual	**Predicted**	
	0	**1**
0	39 (5.5%)	168 (23.7%)
1	8 (1.1%)	493 (70%)

Analysis also reveals that among all our input variables to construct the model, only x, t_x , and T have actually been used in tree construction (in both data mining models) and average monetary value of transactions (m_x) was unable to be a proper 'split' in constructing the decision tree to accurately indicate customer status.

Additionally, further analysis of the cost-sensitive model shows that while T is constant, the churn probability decreases for all 'x', as "time of the last purchase" increases. The only exception is customers with large t_x and small number of repeated purchases; such customers have a higher risk of churn, when compared with clients with the same 'recency' of purchase but with a larger number of repeated purchases (see Figure 2).

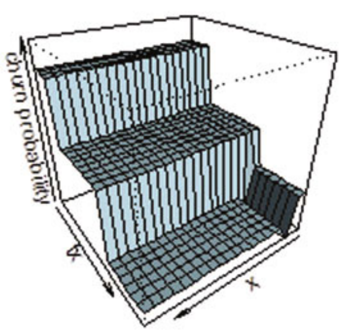

Figure 2 Plot of model's response for the "time of the last repeated purchase" $(\boldsymbol{t_x})$ and "number of repeated purchases" $(\boldsymbol{x})$

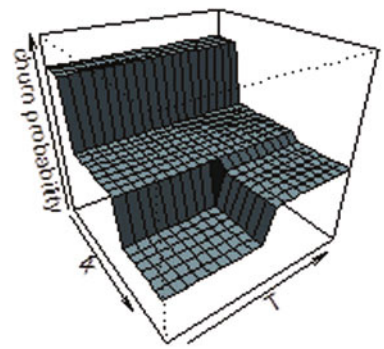

Figure 3 Plot of model's response for the "time of the last repeated purchase" $(\boldsymbol{t_x})$ and "length of observation period" $(\boldsymbol{T})$

On the other hand, when the number of repeated purchases is considered to be fixed, it can be observed that customers with large t_x are more likely to churn when their observation period increases (see Figure 3). In other words, referring to Figure 1, customers whose '0' point is close to 'T' are less likely to churn when the length of $(0, t_x)$ gets larger within the $(0, T)$.

SUMMARY

Comparing the predictive accuracy of the developed model shows that data mining models show almost the same performance as the probability models (see Table 4).

Table 4 Prediction accuracy of the developed models

Model Type	Accuracy, %
Pareto/NBD	76.4
Decision Tree (CART)	75.3
Decision Tree (CART; Cost sensitive learning)	75.1

However, by more closely examining the performance of the models, one notices that while Pareto/NBD and decision tree models have identified 61.1% and 60.6% of churners respectively, the decision tree with cost sensitive learning has been able to identify 70% of churners in our test sample.

CONCLUSION

The current study investigated the performance of existing customer churn modeling approaches. Three predictive models were developed and compared using Pareto/NBD model from probability modeling stream and decision tree from data mining stream. Results revealed that although the Pareto/NBD model shows a slightly better performance in terms of general accuracy, the decision tree model with cost sensitive learning has the upper hand in terms of its ability to identify customers who are likely to churn in a given future time. Obviously, making the same comparison on other datasets, from other sectors, and with more observations and predictors would enable us to better examine the generalizability of our findings.
Furthermore, in the as it was mentioned earlier BG/NBD model, as a simpler version of PARETO/NBD, is unable to predict customers' status for the ones with no repeated purchase in calibration period. This fact made us not to use BG/NBD model

in the current. However, as another direction of research one can use only the subset of those customers who have purchased in the calibration period and by constructing the BG/NBD model on this subset, analyze the performance of BG/NBD model against other models.

REFERENCES

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. New York: Chapman and Hall.

Buckinx, W., & Van den Poel, D. (2005). Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research, 164*(1), 252-268.

Burez, J., & Van den Poel, D. (2008). Separating financial from commercial customer churn: A modeling step towards resolving the conflict between the sales and credit department. *Expert Systems with Applications, 35*(1-2), 497-514.

Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications, 36*(3), 4626-4636.

Colgate, M. R., & Danaher, P. J. (2000). Implementing a customer relationship strategy: The asymmetric impact of poor versus excellent execution. *Journal of the Academy of Marketing Science, 28*(3), 375-387.

Cui, D., & Curry, D. (2005). Prediction in marketing using the support vector machine. *Marketing Science, 24*(4), 595-615.

Ehrenberg, A. S. C. (1959). The pattern of consumer purchases. *Journal of the Royal Statistical Society, 8*(1), 26-41.

Ehrenberg, A. S. C. (1972). *Repeat Buying; Theory and Applications*. Amsterdam: Norht-Holland Publishing Company.

Fader, P. S., & Hardie, B. G. S. (2001). Forecasting repeat sales at CDNOW: A case study. *Interfaces, 31*(3), 94-S107.

Fader, P. S., & Hardie, B. G. S. (2009). Probability models for customer-base analysis. *Journal of Interactive Marketing, 23*(1), 61-69.

Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005). "Counting Your Customers" the easy way: An alternative to the Pareto/NBD model. *Marketing Science, 24*(2), 275-284.

Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2008). Computing P(alive) using the BG/NBD model. Retrieved from http://www.brucehardie.com/notes/021/ website:

Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research, 34*(10), 2902-2917.

Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining*: The MIT press.

Hoppe, D., & Wagner, U. (2007). Customer base analysis: The case for a central variant of the Betageometric/NBD model. *Marketing—Journal of Research and Management, 3*(2), 75-90.

Kamakura, W., Mela, C., Ansari, A., Bodapati, A., Fader, P., Iyengar, R., . . . Wilcox, R. (2005). Choice models and customer relationship management. *Marketing Letters, 16*(3), 279-291.

Kotler, P., & Keller, K. L. (2006). *Marketing Management* (12th ed.). New Jersey: Pearson Prentice Hall.

Lemmens, A., & Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research, 43*(2), 276-286.

Ngai, E., Xiu, L., & Chau, D. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications, 36*(2), 2592-2602.

Olafsson, S., Li, X., & Wu, S. (2008). Operations research and data mining. *European Journal of Operational Research, 187*, 1429-1448.

Rosenberg, L. J., & Czepiel, J. A. (1984). A marketing approach for customer retention. *Journal of Consumer Marketing, 1*(2), 45-51.

Schmittlein, D. C., Morisson, D. G., & Colombo, R. (1987). Counting your customers: Who are they and what will they do next? *Management Science, 33*(1), 1-24.

Wübben, M. (2008). *Analytical CRM; Developing and Maitaining Profitable Customer Relationships in Non-contractual setting* (1st ed.). Frankfurt, Germany: Gabler.

Wübben, M., & Wangenheim, F. V. (2008). Instant customer base analysis: Managerial heuristics often “get it right”. *Journal of Marketing, 72*(3), 82-93.

Xie, Y., Li, X., Ngai, E., & Ying, W. (2009). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications, 36*(3), 5445-5449.