ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 483-488

Data Mining Techniques on the Evaluation of Wireless Churn

Jorge B. Ferreira, Marley Vellasco,
Marco Aurélio Pacheco, Carlos Hall Barbosa

ICA: Computational Intelligence Laboratory
Department of Electrical Engineering, PUC-Rio
R. Marquês de São Vicente 225, Gávea
Rio de Janeiro, 22453-900 RJ, Brazil
(jorge, marley, marco, hall)@ele.puc-rio.br

Abstract. This work focuses on one of the most critical issues to plague the wireless telecommunications industry today: the loss of a valuable subscriber to a competitor, also defined as churn. Analytical methods and models intrinsic to decision technology and machine learning are here evaluated, in an effort to provide the necessary intelligence to identify and understand troublesome customers in order to act upon them before they churn. Making use of a large real-world database, a thorough analysis is performed. First, due attention is given to data representation, with input selection methods being employed in the search of the most relevant attributes. Then, the predictive and explanatory power of four families of models is compared: neural networks, decision trees, genetic algorithms and neuro-fuzzy systems. To conclude, light is shed upon the possible savings and profits resulting from the application of the developed methodology in the retention strategies of wireless carriers.

1. Introduction

Few markets today are as competitive or dynamic as the one modern wireless telecommunications industry inhabits, as [7] well portrays. With mobile technology evolving at a breathtaking pace and the ever growing number of carriers operating simultaneously in the same areas, consumers face a huge array of options from which to choose and satisfy both their hardware and service needs. The amazing flood of advertisement promoted by rival carriers is more than a proof of the rampant battle that is taking place to obtain new clients and, more importantly, to entice customers from the competition to change their hearts. This is where churn arises.

Defined as the act of abandoning a carrier in favor of a competitor, churn has become one of the most crucial business issues confronted today by wireless companies. Given the cost of acquiring a new customer versus the resources needed to maintain an existing one, great competitive advantage and immense financial savings, perhaps as much or more than $50 million, can be achieved [8] if the churn of valuable subscribers can be accurately predicted in time to be acted upon.

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 483-488

Hence, the goal here is to perform accurate classification of future churners while spawning, concurrently, the knowledge that will lead to a better comprehension of the reasons that cause such behavior. In sections 2 and 3, suitable care is given to data representation. Section 4 delineates the methodology utilized in the search for the best churn prediction, with various types of models being evaluated and compared (neural networks, neuro-fuzzy systems, decision trees and genetic algorithms). The analysis of the obtained results and all relevant comparisons are located in section 5. Section 6 encompasses a brief profitability analysis of the generated models, in order to identify those most financially appropriate for wireless carriers to base their decisions upon. Final conclusions and suggestions will be located at Section 7.

2. The Data Set

The data set used throughout this work is from a major wireless carrier operating in Brazil. The database consists of 100,000 subscribers, with the great majority of these being users of a single phone number (the most common type of client available).

Available data refers to nine months of consumer usage, from October 2000 through June 2001. During this period, average total monthly revenue generated by the subscribers in the database was about R$7,290 million and the average revenue per individual was close to R$73.

The definition of the target variable, whether each customer can be considered a churner or a non-churner, was made based on the closing of all services held by a subscriber at the end of the base month. The base month chosen was May 2001, so only clients who cancelled their contracts during the month of June 2001 were tagged as churners. Those who left the company before the base month were not included in the analysis. This procedure for identifying churn is the same as the one currently employed by many carriers in the business.

As a result of such definition, only 1,245 of the subscribers present in the data set were considered churners, corresponding to a 1.245% monthly churn rate for the studied carrier. Unfortunately, this small percentage meant the resulting analysis would tend to non-churners. To circle this problem an oversampling procedure [1] was employed, resulting in a final data set of 3,500 observations, with a proportion of approximately 28% of churners. While a solution, this had a drastic impact on the available set of information, no doubt affecting the overall performance of the models used.

Having concluded the vital steps of defining the target variable and the observations to be used, attention is now focused on the available customer information that hopefully will allow the construction of an accurate predictive model.

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 483-488

In the wireless telecommunications industry several kinds of subscriber information are readily available in massive amounts of corporate data. Using broad categories, information relevant to churn analysis can be summarized as follows:

- *Billing data* (monthly revenue, cost of roaming, etc);
- *Usage data* (air time, types of services used, etc);
- *Customer demographics data* (age, gender, region, etc);
- *Customer relationship data* (rate plan, technology used, handset age, etc);
- *Market data* (competitor rates, advertising costs, etc).

There were a total of 37 different attributes available in the database.

3. Data Representation

The goal at this stage is to judge the explanatory power of each input variable in respect to the act of churning. By wisely choosing inputs that are closely related to the target and improving this relationship through a series of data transformations, more precise models can [8] and will be designed. Also, dimension reduction is quite desirable.

Two different sets of data were assembled in order to portray the advantages of a careful data representation. First, a *simple* representation was constructed, with almost no modification and only minor selection of the inputs being made. Next, for an *enhanced* representation of the data, many methods were employed. Among the most meaningful of them is the application of various input selection algorithms to the data set, in order to identify attributes that better explain the churn behavior of customers. The methods utilized are known as LSE (Least Squares Estimator) and SIE (Single-Input Effectiveness). It can be said that both LSE and SIE, each in its way, try to measure the importance of an input in explaining the variability of a target. All details relative to these input selection procedures can be found at [3] and [10], respectively. Through these methods, it was established that variables related to the pattern of air time consumption by subscribers (such as roaming time, night usage and long distance air time) are decisive in defining churn for the studied database. All numerical variables were normalized by range, both in the *simple* and *enhanced* representations.

4. Methodology

The challenge at hand, as often stated before, is to build a model that can accurately distinguish churners from non-churners. To face this difficult task, four different types of models have been evaluated (relevant references follow each model): MLP Neural Networks [1, 2], C4.5 Decision Trees [1, 9], Hierarchical Neuro-Fuzzy Systems [4] and a data mining tool named Rule Evolver, based on Genetic Algorithms (GA) [6]. MLP neural networks evaluated here had a single hidden layer, were trained through

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 483-488

backpropagation and had the optimal number of hidden neurons determined empirically. The goal is to compare the predictive performance of each kind of model and point in the direction of the one better suited for the problem. Moreover, the improvements achieved by the *enhanced* data representation (10 inputs) over the *simple* representation (with 20 inputs) will be assessed. All results presented in the next section shall refer to the optimized versions of the proposed models, obtained after an extensive process of best parameter estimation and performance evaluation.

To estimate the generalization error of the tested models, a ten-fold cross validation was performed, using the same splits and maintaining churn proportion in every training, validation and test data sets, for all models. Training data corresponded to 70% of the available observations, validation to 20% and test to 10%.

5. Results and Discussion

As a measure of classificatory accuracy, confusion matrices are a traditional way to determine how well a given predictor performs. They clearly illustrate the precision of a classification model, comparing correct predictions and error for each of the existing classes (in this case churners and non-churners). Table 1 portrays the confusion matrices for the test data set (350 observations) of all models to be compared, bearing the results of the *simple* and *enhanced* representation for each model.

Table 1. Confusion matrices for the test data set of each model

		***Simple* Representation**		***Enhanced* Representation**	
		Predicted		**Predicted**	
Model	**Actual**	non-churner	churner	non-churner	churner
Neural Network	non-churner	70 %	30 %	77%	23%
	churner	35 %	65 %	27%	73%
Decision Tree	non-churner	69 %	31%	70%	30%
	churner	40 %	60 %	33%	67%
Neuro Fuzzy	non-churner	58%	42%	69%	31%
	churner	33%	67%	25%	75%
(GA) Rule Evolver	non-churner	44%	56%	52%	48%
	churner	36%	64%	32%	68%

According to Table 1, the *enhanced* models significantly outperform their *simple* counterparts, attesting for the vital importance of a careful data representation. Among the models, neural networks with 15 hidden units accomplished the best classification, followed somewhat closely by the neuro-fuzzy system and the decision tree. The genetic algorithm based model proved to be ill suited for the data set at hand.

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 483-488

At this point, it is vital to notice that the oversampling procedure that was required to adequate the data set produced a small number of observations (only 3,500), a fact that undoubtedly contributed for worse results than could have been obtained, for example, with a data set of 50,000 subscribers containing 15,000 churners.

6. Profitability

Confusion matrices are a nice way to measure the predictive power of the proposed models. However, they fail to demonstrate the full potential for savings and profits such predictors can lend to an enterprise actively using them as a marketing and strategic tool [5, 7]. To achieve this, a methodology identical to the one proposed by [8] will be utilized, describing in detail the cost savings born on successful churn reduction strategies based on the predictive accuracy of analytic models (ultimately the gap between the cost of an incentive and the cost of the loss of a profitable customer).

Table 2 shows the cost savings accrued for the *enhanced* representation by each model for two scenarios: one equal to the carrier's situation in 2001 (monthly churn rate of 1.25%) and another reflecting a higher churn rate of 3%, more realistic for the current market situation in 2003, with four carriers operating simultaneously instead of two. In all estimations, it is assumed that 50% of the possible churners offered an incentive will be retained, that the average monthly bill of subscribers is $80, that the cost of an incentive is $25 per person and that only 20% of those predicted as churners can be contacted in a given month. Results are for a carrier with a total of 2 million subscribers.

Table 2. Savings accrued for the *enhanced* representation by each model

	Monthly Churn Rate of 1.25%		**Monthly Churn Rate of 3%**	
Model	**Monthly Savings per Subscriber**	**Total Annual Savings**	**Monthly Savings per Subscriber**	**Total Annual Savings**
Neural Network	$44.20	$13,260,000	$97.87	$70,464,000
Decision Tree	$5.45	$1,635,000	$75.45	$54,324,000
Neuro Fuzzy	$11.30	$4,890,000	$88.63	$63,816,000
(GA) Rule Evolver	($63.80)	($19,140,000)	$48.20	$34,704,000

As Table 2 depicts, the neural network model with *enhanced* data representation was able to provide expected savings of $44.20 per "churnable" client. This translates into more than $15 million a year, and that's only because the churn rate in the analyzed carrier was quite low (1.25%). If the same model were to be employed by a carrier with a 3% monthly churn rate, expected savings would rise to an incredible figure of $70 million a year.

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 483-488

7. Conclusions

It was shown that solid analytical modeling can add precious value to churn management strategies in the Brazilian wireless telecommunications industry. It was proved that great cost savings can be drawn from well-founded churn retention actions. Through input selection, it was verified that the most important variables in defining churn for the studied database are those related to the pattern of air time consumption by subscribers, like roaming time, night usage and long distance air time. This proves that, along with providing knowledge about the most crucial modeling variables, input selection methods can grant noteworthy intelligence over which the key factors behind churn are.

An important aspect to be explored in future works would be the validation of the data preparation and modeling procedures with other data sets, particularly ones with a larger monthly churn rate, a larger number of observations and more available inputs than the one explored here.

References

1. BERRY, Michael J. A., LINOFF Gordon: Mastering Data Mining: The Art and Science of Customer Relationship Management. John Wiley & Sons (2000).
2. BISHOP Christopher M.: Neural Networks for Pattern Recognition. Oxford University Press (1996).
3. CHUNG, Fu-Lai, DUAN, Ji-Cheng: On Multistage Fuzzy Neural Network Modeling. IEEE Transactions on Fuzzy Systems. Vol 8, pp 125-816 (2000).
4. GONÇALVES, Laércio B.: Modelos Neuro-Fuzzy Hierárquicos BSP para Classificação de Padrôes e Extração e Regras Fuzzy em Bancos de Dados. MSc Dissertation, Departamento de Engenharia Elétrica, PUC-Rio (2001).
5. LITTLE, John D.C.: Decision Support Systems for Marketing Decisions. Journal of Marketing, 43 Summer (1979).
6. LOPES, Carlos H.: Classificação de Registros em Bancos de Dados por Evolução de Regras de Associação Utilizando Algoritmos Genéticos. MSc Dissertation Departamento de Engenharia Elétrica, PUC-Rio (1999).
7. MATTISON, Rob: Telecom Churn Management. APDG Publishing (2001).
8. MOZER, Michael C. et al: Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry. IEEE Transactions on Neural Networks, Vol. 11, No. 3 (2000).
9. QUINLAN, J. R.: Simplifying Decision Trees. International Journal of Machine Learning Studies, vol. 27, pp 221-234 (1987).
10. YI, Cao, DIANE, Rossiter: An Input Pre-screening Technique for Control Structure Selection. Computers Chem. Engng. Vol 21, No 6, pp 563-5369 (1997).