

Predicting the Credit Limits of Subscribers to Mobile Communication Services*

Falk Wagner
Chair of Mobile Commerce and Multilateral Security
Frankfurt University, Germany

eMail: falk.wagner@m-lehrstuhl.de

Abstract

Mobile communication has become a matter of course for many people. As markets saturate, the care and retention of existing customers becomes a key element for revenue stabilization for Mobile Communication Network Operators. We present a predictive data mining model to reduce the rate of forced churn as a consequence of non-payment: Estimations of subscribers' open amounts if being payers or non-payers allow to prevent subscribers from overspending – and ultimately churning – thus prolonging the customer relationship dwell time and securing future revenues.

Introduction

Mobile communication has become a matter of course for many people. Current subscription coverage rates of central European countries, however, indicate saturated markets, leading to intensified competition among operators. As margins from operations decrease, operators will have to review their practice of subscriber acquisition through ever-increasing dealer commissions and handset subsidies. Improved customer management may instead shift to the focus in order to maximize revenue from current subscribers.

While the prevention of voluntary subscriber churn – customers leaving after the minimum contract term or during that period without debt – has been on operators' agenda for a long time, prevention of forced churn – usually due to default of payment by the subscribers – has often not been in the key focus. This is striking as fixed-line to mobile convergence and new high-priced data services are expected to drive up subscriber spending, a development likely to increase the probability of budget exceedance and subsequently default of payment by subscribers.

Unfortunately, it is very difficult to provide accurate estimation of a subscriber's solvency limit from existing credit scores where in place: This is due to the fact that churn motives such as low solvency – putting limits on the payment of unexpectedly high bills – and impulsive service usage – leading to the exceedance of a subscriber's planned budget – cannot be distinguished within a score class. To establish the missing link between churn predictors and the usage pattern actually initiating the churn event, a predictive data mining model was constructed.

The remainder of the paper is organized as follows: Chapter 1 gives a brief overview on credit scoring then presents our approach. Chapter 2 provides an introduction to data mining and to algorithms for predictive modelling. Chapters 3, 4 and 5 describe the data set available, the configuration and training of an Artificial Neural Network (ANN) and the prediction accuracy achieved. Chapter 6 concludes our findings and provides an outlook to further research.

* The author would like to thank Deutsche Telekom Business Academy for funding of his research and express his gratitude to T-Mobile Germany departments F14 (Mrs. Anke Heuermann, Mr. Pius Jakobs) and TP&E-12D (Mr. Martin Rauh) for their support.

Prediction of Customer Default

Assessment of customers' risk of default is crucial to many industries. Fitzgerald (1932) and Durand (1941) were the first to investigate the predictive power of lender characteristics for the probability of default of corporate and consumer credit respectively. Credit scores present a quantitative measure of the predicted probability of default and are used to control customer acquisition, credit limit adjustments and collection processes in various scenarios.

Credit scoring models evaluate the likely membership of customers to groups of payers and non-payers. The models' classifications can be understood as the outcome of a performance period (denoted by t) that is a function of the individuals' characteristics at the beginning of that period (denoted by $t-1$). In this context the outcome is regarded unconditional of the behaviour within the performance period – default or non-default is intrinsic to the individual i 's characteristics in $t-1$. To formalize we designate this characteristics $C_{i(t-1)}$ and the period's outcome – default or non-default – as states $S_{i(t)}$:

$$(1) \quad S_{i(t)} = F(C_{i(t-1)})$$

By including account, usage and payment data ($P_{i(t-1, \dots, t-n)}$) from previous billing periods into the predictor set, credit or application scoring is extended to behaviour scoring:

$$(2) \quad S_{i(t)} = F(C_{i(t-1)}, P_{i(t-1, \dots, t-n)})$$

Both approaches seem appropriate when the factors influencing $S_{i(t)}$ are fixed or capped during the performance period, e.g. by agreed loan redemptions or credit limits. However in setups with no such ceiling amounts, the open amount as a likely determinant of $S_{i(t)}$ is variable. In our model we assume that default on payment is a consequence of open amounts exceeding the liquid assets of a customer. This implies that the outcome of the performance period is a result of the behaviour within this period and the associated open amounts $OA_{i(t)}$:

$$(3) \quad S_{i(t)} = F(C_{i(t-1)}, P_{i(t-1, \dots, t-n)}, OA_{i(t)})$$

An important implication of this approach is that default and non-default are states that may be adopted by every customer depending on the total of open amounts he faces. It further presumes that default on payment can be prevented through control of customer spending: Through well timed suspension of services, customers may be prevented to overspend and exceed their liquid assets. Obviously efficient control is often hindered by interactions of the customer with third parties or sudden changes of life circumstances, such as unemployment or alimony obligations, unobservable to business partners. For the sake of simplicity, we assume that a particular business relationship is (only) accountable for the insolvency event if the customer's spending on it has considerably exceeded previous spending levels.

The above considerations imply the existence of open amounts that cause the customer to change from the state of non-default to default. Since we intend to actually forecast these state-conditioning amounts, we rewrite formula (3) as:

$$(4) \quad OA_{i(t)} = F(C_{i(t-1)}, P_{i(t-1, \dots, t-n)}, S_{i(t)})$$

Different from credit and behaviour scoring, the period's outcome of (non-)default of a customer is no longer a target but a predictor variable. Its inclusion in the predictor data set allows the calculation of the open amounts associated with each state through the manipulation of $S_{i(t)}$: The values correspond to the service usage of the customer if he does or does not default in the next billing cycle, respectively. The limit of solvency of the customer lies in the range between those amounts which we aim to estimate based on techniques for predictive data mining.

Predictive Data Mining for Regression

The term *Data Mining* is used in various contexts and diverse definitions can be found throughout literature. The essence of all definitions is the support to the extraction of information from data through model building (Berry and Linoff 2004, pp. 36-42; Hormozi and Giles 2004). The construction

of models based on information inferred from examples is called *Inductive Learning*, its automation is the subject of *Machine Learning*. (Holsheimer and Siebes 1994, pp. 7-24) define Data Mining as a sub-discipline of Machine Learning where samples are retrieved from databases. This concept is congruent with Fayyad, Piatetsky-Shapiro et al. (1996, p. 28) who define Data Mining as the step of the *Knowledge Discovery in Database* (KDD) process where algorithms are applied to extract models from data.

The classifications of data mining tasks vary in literature (Hormozi and Giles 2004); However there is an agreement on the prominent role of predictive modelling. This category comprehends estimations of categorical (*Classification*) and quantitative variables (*Regression, Estimation*) based on a set of predictor variables (Hand, Manilla et al. 2001, pp. 11-15). Well suited algorithms for estimation are linear and logistic regression, ANN, radial basis functions, k nearest neighbours and k means (Chen 2001, p. 129). The latter three algorithms are clustering techniques designed for the segmenting of heterogeneous population into a number of more homogeneous subgroups. Such undirected data mining methods can provide estimations of a target variable based on the target's values within the subgroups.

Previous credit scoring projects based on sample sets similar to the one used for the construction of our model have revealed nonlinear relationships and interactions between variables – for instance subscriber age and purchasing power of ZIP areas. Due to their robustness in this respect (Liu 2003, p. 67) we decided to use ANNs for our estimation model. Often cited drawbacks such as the non-interpretability of neuron interconnections had to defer to their beneficial features.

The development of the estimation model was carried out in three steps. Firstly the data set was examined and predictor variables were selected and created where reasonable. Secondly an ANN was trained to estimate the subscriber's state-dependent open amount for the upcoming billing cycle. Thirdly the model performance was tested on a hold-out sample and its accuracy was assessed.

The Data Set

The data set comprised anonymized details of individuals that subscribed to a German mobile communication service provider (SP) in summer 2004. For these customers master file, invoice, and aggregated usage data were available. Based on the subscribers' account status we were able to identify the invoices that for some subscribers had initiated the dunning process and ultimately led to the cancellation of the subscriber's contract by the SP (*Forced churn*). Those invoices were marked with a *churn flag*.

The aggregated usage data included the amounts subscribers had spent per service (calls to fixed-line numbers, short message services (SMS), roaming etc.) and billing period. Exceptional changes in usage, for instance sharp usage declines indicating an eventual "Bill Shock", were of great interest to us, particularly if accompanied by dishonoured direct debits. These in-depth investigations were to highlight data constellations that may indicate subscribers' budget for mobile communication services.

Master file data, usage data of three consecutive billing periods and the "exceptional event" flags were combined in a predictor sample. We also included the churn flag that indicated if the next periods' invoice was going to initiate the churn. The corresponding invoice amount was saved as the target variable the model shall estimate.

The sampling routine had two drawbacks: First, it took usage pattern out of their seasonal context. With respect to roaming and international telephony we wished to include a variable that could describe the important seasonal fluctuations ex-ante. A number of statistical variables were tested. The distribution of cross-border air traffic and tourist overnight-stays during the year were found to be good predictors and were included in the set of predictors. Secondly we found the sample to include relatively few churn events, since ANNs are sensitive to class proportions we over-sampled and weighted both groups equally.

Configuration and Training of the ANN

The development of ANNs has been provoked by research on the representation of events in the human nervous system (McCulloch and Pitts 1943). To model the information processing capabilities of these systems, networks of interconnected artificial neurons are constructed and trained in order to reproduce target values based on data presented to the network's input layer.

The training of a network consists of the iterative adjustments of the weights of links between the neurons with the goal of minimizing the error between target values and network outputs. The extent of the adjustment per training epoch is defined by the training learning rate; its allocation among the links is directed by training algorithms. These estimate the direction of steepest error descent on the error surface spanned by the possible combinations of the link weights.

For our modelling task we choose a three-layer perceptron network with back-propagation. The neuron transfer functions were sigmoid in the hidden and linear in the output layer. To speed up the convergence of the network to the optimal weight combination, a training function with gradient descent momentum and adaptive learning rate was employed. The degree of convergence depends also on the number of training epochs for weight adjustment. This setting is sensitive since the network may overfit to the training data while losing its prediction accuracy on unseen data. To assess the risk of overfitting, we cross-validated the learning performance on the training and the hold-out sample: It could be found that within 5000 training epochs the network did not overfit the training sample – any increase in performance on the training sample also increased the performance on the hold-out sample (Figure 1). This is an indicator that the selection of the samples from the basic population happened randomly and that both samples are representative of the overall population.

Repeated trainings revealed that important performance gains were realized within the first 1000 training epochs (figure 2). To reduce computational costs we limited the training sessions to 1000 epochs.

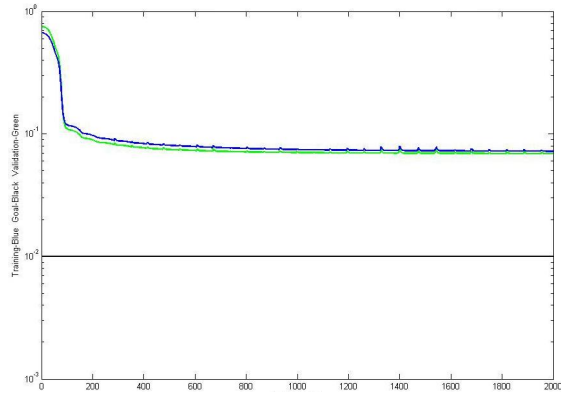


Figure 1: Network performance on training and validation sample per training epoch

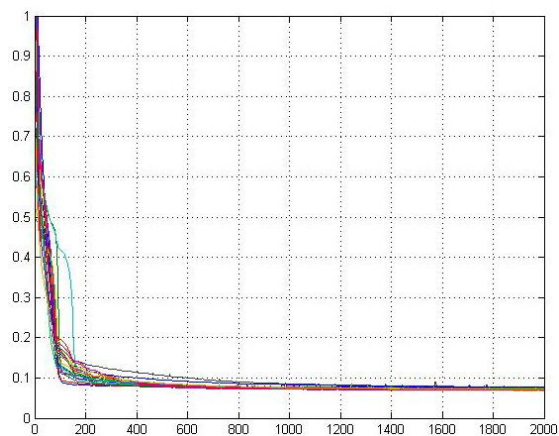


Figure 2: Performance of different network configurations on training sample per training epoch.

To select the appropriate ANN configuration for our estimation task, the training performance was repeatedly tested for several configurations that varied in the number of neurons in the hidden layer. Based on a comparison of the configurations' root mean squared error as the average mismatch between target values and networks' outputs, the best performing configuration was selected for the evaluation of the model performance.

The Model Performance

The performance of the model was evaluated under two aspects. First, the model serves as an estimation function of the usage of mobile communication services by subscribers depending on their state in the upcoming billing cycle. Second, based

on the model, credit control may suspend subscribers when reaching certain open amounts to prevent default on payment from overspending. Our intention was to assess the “hit rate” of the model in such a scenario.

With respect to its estimation accuracy, figure 3 shows that the model predicts values all over the range of open amounts for both churners and non-churners. This is an important characteristic as it indicates that the model does not simply assign subscribers to separate bounds of values depending on their churn flags but discovers systematic interactions of the diverse subscriber characteristics. As a result estimation values for high-spending non-churners and low-spending churners may overlap.

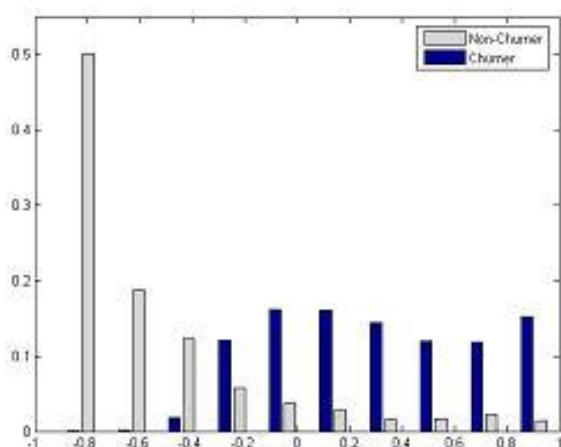


Figure 3: Histogram of the network output

Figure 4 visualizes the distribution of the squared errors of the network output to assess the overall estimation accuracy of the model: Its strong orientation to the left indicates small prediction errors on the majority of estimated cases.

We further manipulated the churn flag in the predictor sample to simulate the hit rate of the model when used for the setup of credit limits: subscribers were assigned to churner and non-churner state respectively and tests were run for both constellations. The churner invoice amount was found to be higher than the non-churner invoice amount for all subscribers – this result is congruent with our reasoning in chapter 2.

Two estimations were obtained as results of our “sensitivity analysis”: The open amount

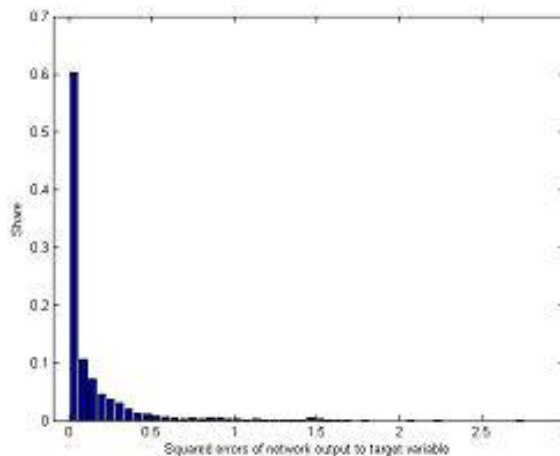


Figure 4: Histogram of squared network errors.

expected for the subscriber if being a non-churner and the open amount for the subscriber being a churner in the next period (both values being subject to the estimation errors as outlined above). Since we assumed the invoice amount to drive the churn event, the SP is able to prevent the churn by suspending subscribers at their credit limit and preventing them from overspending.

From a financial perspective, the determination of the optimal credit limit within the estimation interval corresponds to a trade-off between immediate, but insecure since “above the limit”, revenues and the net present value of expected, “secure” since credit limit controlled, revenues from the residual (eventually extended) customer dwell time. The projection of subscriber revenues goes beyond the focus of this paper and is subject to further research. For the sake of simplicity the subscriber solvency limit as the optimal credit limit is assumed to be located at the centre of the estimation interval. In this context a subscriber is classified as a non-churner if the observed open amount is closer to the non-churner than to the churner estimation.

Predicted Actual	Non-Churn	Churn	Total
Non-Churn	46.2 %	3.9 %	50.2 %
Churn	14.9 %	34.9 %	49.8 %

Table 1: Confusion matrix of hit rate at credit limit

The confusion matrix in table 1 presents the share of subscribers (in)correctly classified based on the credit limit. From a modelling perspective, misclassifications are due to uncommon high-spending behaviour of non-churners and unexpectedly low open amounts of churners. Chances are that additional master file data and usage flags refine the model of understanding subscribers' solvency. The acceptability of the hit rates depends on the costs associated with misclassifications, namely lost revenues from falsely suspended non-churners and bad debt losses from non-suspended churners.

Discussion and Outlook

The described model addresses the prevention of forced churn of subscribers to mobile communication services due to non-payment. In contrast to classification approaches to predict the association of a subscriber to groups of (non-)churners, the model estimates at which open amount a non-churner becomes a churner.

By manipulation of the churn flag prediction variable, one can query the state-associated open amounts from the model to define subscriber-individual credit limits. A trial of a credit limit set to the centre of the estimation interval proved a good hit rate. The sophisticated determination of the solvency limit within the identified range of open amounts is subject to further research. Future work should also benchmark the accuracy of the ANN approach against alternative regression techniques. A well-performing regression technique is of particular interest to profit scoring.

Apart from future improvements, the model at its current state represents a sophisticated but robust prototype for data-driven determination of credit limits. The benefits envisaged include increased revenues from subscribers prevented from churning as well as improved customer satisfaction from reduced false alarms of legitimately "high-spending" subscribers.

References

Berry, M. J. A. and G. Linoff (2004). *Data Mining techniques: for marketing, sales and customer relationship management*. Indianapolis, Indiana, Wiley Publishing Inc.

Chen, Z. (2001). *Data mining and uncertain reasoning: an integrated approach*. New York, John Wiley & Sons.

Hand, D. J., H. Manilla, et al. (2001). *Principles of Data Mining*. Cambridge, MA, The MIT Press.

Holsheimer, M. and A. Siebes (1994). *Data Mining - The Search for Knowledge in Databases*. Amsterdam, The Netherlands, Centrum voor Wiskunde en Informatica (CWI).

Hormozi, A. M. and S. Giles (2004). "Data Mining: A Competitive Weapon for Banking and Retail Industries." *Information Systems Management* 21(2): 62-71.

Liu, Y. (2003). *A theoretical and empirical study on the data mining process for credit scoring*. Faculty of Economics and Business Administration. Göttingen, Georg-August University.

McCulloch, W. S. and W. Pitts (1943). "A logical calculus of the ideas immanent in nervous activity." *Bulletin of Mathematical Biophysics* 5: 115-133.