Applying data mining in telecommunications

Radosavljevik, D.

Chapter 4

Preventing Churn in Telecommunications: The Forgotten Network

Radosavljevik, D., van der Putten, P.

Published in International Symposium on Intelligent Data Analysis, Springer Berlin Heidelberg, pp. 357–368 (2013).

This chapter outlines an approach developed as a part of a company-wide churn management initiative within T-Mobile Netherlands. We are focusing on an explanatory churn model for the postpaid segment, assuming that the mobile telecom network, the key resource of operators, is also a churn driver in case it under-delivers to customers' expectations. Typically, insights generated by churn models are deployed in marketing campaigns; our model's insights are used for network optimization in order to remove the key network related churn drivers and therefore prevent churn, rather than cure it. The insights generated by the model have caused a paradigm shift in managing the network of T-Mobile Netherlands.

4.1 Introduction

The phenomenon of churn, which denotes loss of a client to competitors, is a key problem across industries. New customers are difficult to find, especially in saturated markets, such as the European mobile communications market. Furthermore, it is far less expensive to retain existing customers than to acquire new ones. Retention is usually a process that identifies customers that are likely to churn, using various predictive modeling techniques, followed by approaching these customers with suitable offers that would persuade the customer into extending the contract. But, can the customer be prevented from even wanting to churn? Can the main churn drivers be mitigated beforehand?

This chapter is focused on a company-wide churn reduction initiative conducted in T-Mobile Netherlands. As explained above, churn/customer retention is typically a marketing based process. But, despite involving predictive analytics, this process is in its nature reactive, because the customer has already decided to churn and an action is being taken to stop this.

In this research we are taking a completely different approach: the model generated here is not to be used for campaigning. Our method attempts to tackle churn by identifying the key reasons why customers decide to churn in order to alleviate them, rather than identify prospective churners. Hence, the research question in this chapter is:

As a different method for model deployment, can a churn model be used to prevent churn by explaining its causes as opposed to using the predictions for targeting customers?

This approach is even more justified taking into account the current and future stringent European Data Privacy regulations, which limit operators' use of customer data for campaigning purposes. This is especially the case with Internet usage data.

The mobile telecommunications network is a key resource for telecom operators. It is the means of service delivery as well as the most frequent touch point with customers. Problems with ability to use the network (services) have been identified by surveys internal to the company, as well as in literature (Section 4.2), as one of the key reasons to churn. But, most of the time, customers are not experts and cannot pinpoint what exactly is going wrong. Most of the time, this is generalized

as "coverage problems". This research is taking a deep dive into various network problems and their relation to customer churn. The main objective here is to identify the problems that customers who have churned were experiencing, so that they can be corrected for the current customer base and reduce their likelihood of churn. In other words, rather than treating symptoms, we are treating the cause of the disease. This research and its outcome have caused a paradigm shift in managing the network with the operator where the research was conducted.

As opposed to chapters 2 and 3, in this research we are focusing on the postpaid customer segment. Even though these customers are bound by contract, which makes the task of churn prediction slightly less challenging, the revenues that are typically generated here are much higher than in the prepaid segment. Furthermore, postpaid customers' service usage is much higher, compared to the prepaid segment; therefore they would be more prone to experiencing network related issues which can potentially lead to churn. The combination of higher usage and revenues makes it easier to justify the network investments needed to remedy their problems.

The rest of the chapter is structured as follows. Section 4.2 describes the related work on telecom churn. Section 4.3 discusses the data set and methodology we used. Section 4.4 contains the results, their application. Limitations and future work are discussed in Section 4.5. Finally, we present our conclusions in Section 4.6.

4.2 Telecom Churn in Literature

Churn in various industries has been a growing topic of research for the last 15 years (Verbeke et al., 2011). According to Ballings and Van den Poel (2012), churn management consists of predicting which customers are going to churn and evaluating which action is most effective in retaining these customers. Retention strategies are in the focus of Hung et al. (2006). However, most often churn prediction and improving model performance were analyzed following one of these two strategies: adding/improving the data to mine and inventing new algorithms or improving the existing ones (Ballings and Van den Poel, 2012).

The remark above is certainly valid in the case of telecom churn literature. Many papers were trying to find the best algorithm that would outperform all others. Multiple works have examined Logistic Regression, Decision Trees, Neural Networks, evolutionary learning, discriminant analysis and Bayesian approaches (Au et al., 2003; Mozer et al., 1999; Neslin et al., 2006; Wei and Chiu, 2002). Other papers analyzed Support Vector Machines, Random Forest, Rotation Forest, Bagging and Boosting (Archaux et al., 2004; Idris, Khan and Lee, 2013; Lemmens and Croux, 2006; Verbeke et al., 2011). In our view, the value of this research is somewhat limited, at least for real world data mining, given the No Free Lunch theorem (Wolpert and Macready, 1997). In the period after 2008, an overwhelming theme in (telecom) churn research was Social Networks Analysis (SNA), claiming to largely improve on existing churn models (Dasgupta et al., 2008; Motahari et al., 2012; Nanavati et al.,

2006; Richter et al., 2010; Wang et al., 2010; Ngonmang, Viennet and Tchuente, 2012; Polepally and Mohan, 2012; Saravanan and Raajaa, 2012). However, as explained in chapter 3, this claim is not generally applicable, at least not in prepaid churn prediction on a European market. Most of the SNA research focuses on the Asian or US Markets.

Taking into account the data perspective, most of the literature, especially the one focusing on SNA, was using features extracted from Call Detail Records (CDRs). Contractual, demographic, billing, handset, customer service, market (competitor's offers), and customer survey data is often used in addition to CDRs (Archaux et al., 2004; Au et al., 2003; Hung et al., 2006; Mozer et al., 1999; Neslin et al., 2006; Wei and Chiu, 2002). Just a few of these papers take into account any network usage related problems as possible factors affecting churn. For instance, dropped calls were considered as potential churn influencers (Ahn, Han and Lee, 2006; Mozer et al., 1999). Service quality in general and innovativeness were marked as churn detractors by Malhotra and Malhotra (2013).

Predictive models trying to explain churn have not received as much attention in literature (Ballings and Van den Poel, 2012; Lima et al., 2009). Nevertheless, there are studies in industries other than telecom illustrating the need to gain insight into causes of churn (Anil Kumar and Ravi, 2008; Buckinx and Van den Poel, 2005). Furthermore, research based on customer surveys claimed that network coverage, mobile signal strength and voice call drops are reasons for customers to churn (Ahn et al., 2006; Birke and Swann, 2006; Malhotra and Malhotra, 2013; Min and Wan, 2009; Seo, Ranganathan and Babad, 2008). However, all these papers were based on survey data, thus perception of quality, not actual network measurements.

It is apparent that in most recent telecom churn research the physical telecom network, which is the means of delivering telecom services, has been largely neglected. At best, the (lack of) quality of voice call usage is considered. To the best of our knowledge, there is little or no research on how Internet usage on a mobile network and its quality parameters might affect churn. This is one of the key reasons why the topic of our research is an explanatory churn model for telecommunications with actual network quality usage parameters at its focus, instead of just the customers' perception of network quality. In addition, this model, unlike the models from related work, is not meant for retention campaigns; instead, it is concentrating on eliminating what we see as one of the crucial causes of telecom churn: poor experience using the services on the network.

4.3 Data Set and Methodology

In this section we will describe the process and the data set used in this research. As mentioned previously, this research was not started with retention campaigns in mind. It was a part of a cross departmental company-wide churn tackling initiative, executed in parallel with regular churn campaigns. Therefore, the objective of this

Table 4.1: List of contractual, demographic and CDR based features

Contractual and demographic features	Features Extracted from CDRs
Contract expiry	Amount of Voice Calls, SMS and
List of services/ products used	Internet Volume (MB) used,
Subscription fee	both local and roaming
Monthly Bill for each of services	Breakdowns of Voice Calls and SMS
Age, gender, zip code	onto national-international,
Handset	internal-external(competitors' network)

research was not to compete with churn models created for campaigning, but to detect whether there are telecom network quality related factors influencing churn and identify potential remedies.

4.3.1 Data Set

The results presented here are based on a random sample of 150,000 consumer post-paid subscribers of the operator from September 2012. This is just a fraction of the overall base. There was a limitation enforced on the data set related to contract expiry date: the sample was limited to subscribers whose contracts were expiring in three months or have already expired; thus only customers at risk of churn were taken into account. Churn was measured for the following two months, October and November 2012, combined.

The final data set consisted of 750 features, gathered by merging tables from CRM and Network databases. In addition to the attributes similar to what was described in Section 4.2 (see Table 4.1), we added features stemming from the CEM (Customers Experience Management) Framework (Radosavljevik et al., 2010*a*) we designed in Chapter 2 (see Figure 2.1), most importantly the Network quality or usability features (see Table 4.2). The features extracted from CDRs and the network quality features represent monthly aggregates. We also examined their respective three-month aggregates, as well as if there is a rising or declining trend in the past three months for any of these features and used these as potential predictors of churn.

4.3.2 Methodology

Our research setup is similar to what we have described in chapter 2. The data originally residing in various CRM and Network quality databases was collected into a single Oracle database (Oracle, 2011), which allowed easier manipulation and data cleansing. For Data analysis, Predictor Selection and Model Development and Assessment we used the commercial tool Predictive Analytics Director (Pegasystems, 2008).

Table 4.2: List of network quality features per category

General Network Quality	Voice and SMS quality	Internet quality
2G and 3G Coverage at home	Count of Dropped Voice Calls and SMS	3G and 2G Data Attempts
	Voice Call Setup Failures	3G and 2G Data Errors
Provisioning Errors	Voice Call and SMS drop rate	3G and 2G Success Rate
	Voice Call Setup Duration (Maximum and Average)	Ratio of 3G usage vs. 2G usage

We divided the sample into training, validation and testing set using the ratio 50:25:25. The validation set was used during the data analysis stage as a "pre-test" set, in order to verify the univariate performance of each predicting variable with relation to churn, established on the training set.

The performance measure used to evaluate the performance of each individual predictor, as well as the models, was Coefficient of Concordance (CoC). As explained in chapter 2, CoC is a rank correlation measure related to Kendall's tau, suitable for evaluating scoring models (Kendall, 1938; Pegasystems, 2008). It is a measure equivalent to the Area under the ROC (AUC). One interpretation of the CoC measure is that in a scoring model it gives the probability that a randomly chosen positive case will get a higher score than a randomly chosen negative case. The CoC value ranges from 50 to 100. Random choice has a CoC value of 50.

All models developed are scoring models, i.e. we calculated probabilities that someone will churn, without setting a cutoff point. As mentioned above, these models were not to be used for campaigning, but for network improvements, therefore setting a cutoff point to strictly classify whether an instance is a churner or not was not necessary. For this reason, using measures such as recall and precision were not applicable in our case.

During the data analysis stage, the continuous variables were discretized into bins. Bins without significant performance difference are then grouped together. Basically, this is a supervised, bottom-up approach to discretization of continuous variables. One of the advantages of this approach is that it can address non-linear effects of variables onto churn: namely, each separate bin got a score which is concordant to churn and this score was used for modeling. This process is similar for symbolic variables. Variables can be inspected via histograms and the discretization settings can be manually changed if deemed necessary. The next step in the process is predictor grouping which assists feature selection. Namely, variables that are correlated to each other are grouped together. A given predictor may have a high univariate performance, but also be correlated with other candidate predictors that are even stronger, hence not adding value to a model. We first used the best predictor of each group and then selected/deselected variables manually to develop the models

Table 4.3: Model Performance

Model Description	Number of Predictors	Performance on Training set (CoC)	Performance on Test set (CoC)
Campaign	3	76.0	75.9
Campaign_PlusNetwork	6	76.8	76.7
ContractEnd_PlusNetwork	5	75.1	74.7
Campaign_MinusContractEnd	5	68.7	68.1
PurelyNetworkBased	5	66.6	66.5

with a good performance, but also good explanatory value.

As explained previously, the topic of this research is not finding the next best algorithm. That is why we used standard algorithms, such as Logistic Regression and Decision Trees based on the CHAID splitting method (Witten and Frank, 2005). These methods also perfectly fit the explanatory nature of our research, because they are easy to interpret. This is an advantage in commercial settings, where people that need to make investment decisions based on the model and implement its results are not data miners.

The modeling process resulted in scoring models: each instance is allocated a rank score concordant with the probability of being a churner. The CoC (AUC) measure was used to measure model quality. In addition, we use gain charts as visual representation of model performance. On the y-axis, these charts show the captured proportion of the desired class (i.e. churners in selection divided by total number of churners) with increasing selection sizes (x-axis, from highest scoring to lowest scoring) (see Figure 4.1).

4.4 Results, Application and Discussion

Even though optimizing model performance is not the topic of this research, we deem it necessary to benchmark our network against the campaigning model. The performance (CoC) of the models we created is presented on Table 4.3.

It is worthwhile mentioning that all models presented here were built using Decision Trees with CHAID splitting criterion, which have an inherent characteristic of dealing with non-linear data. We also tested models using Logistic Regression, but they had somewhat worse performance (0.5 CoC points). Please note that due to the discretization process described in Subsection 4.3.2, this implementation of logistic regression is able to handle non-linear dependencies too. Furthermore, in order to test for non-linear interaction effects between a combination of two variables and churn we created close to 280,000 new predictors using two way combinations of all of the 750 variables. However, no strong non-linear effects were noted.

As can be seen on Table 4.3, adding network related features to a campaigning

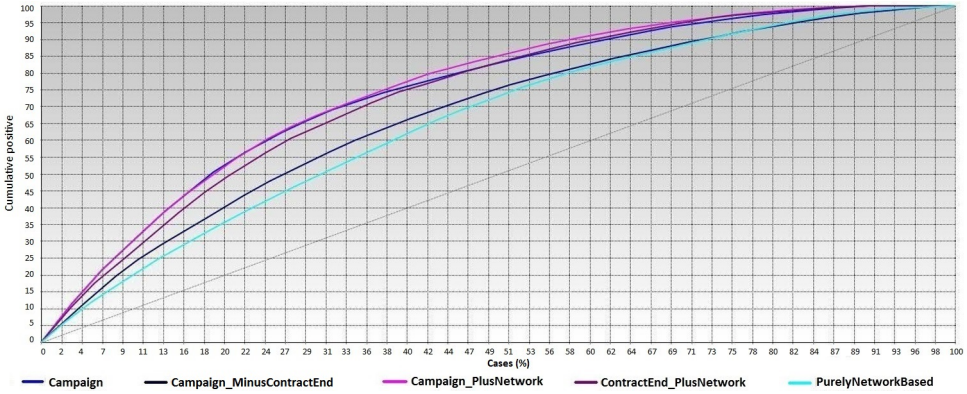

Figure 4.1: Gain Charts of Models Used

model (model Campaign_PlusNetwork) only marginally increases performance (1 CoC point), visible on Figure 4.1 only after the 40th percentile of cases ranked by churn, which confirms our result from (Radosavljevik et al., 2010*a*). However, campaigning wise, this has no meaning because rarely do campaigns address more than 40% of the base that is at churn risk.

The PurelyNetworkBased model, which is the topic of our research, had the weakest performance. Nevertheless, just for comparison reasons, we built a Campaign model without the strongest predictor - Contract End (Campaign_MinusContractEnd) and a model based on a combination of just the Contract End and Network Factors (ContractEnd_PlusNetwork). The Campaign_MinusContractEnd model performed only somewhat better than the Pure Network model (1.5 CoC on the test set in Table 4.3, or 5% more churners in the Top 20% of the scores on Figure 4.1), and the model ContractEnd_PlusNetwork performed only marginally worse than the campaigning model (1.3 CoC on the test set in Table 4.3, or 4% less churners in the Top 20% of the scores on Figure 4.1). The conclusion here is that, less the Contract End variable, the network quality parameters from our Purely Network Based model performed nearly as well as the other predictors.

However, performance was not the main topic of our research. The main aim was the explanatory value of our model. On Figure 4.1 it is shown that Purely Network Based Model could address the 35% of churners in the top 20% of scores, while the Campaign model addressed nearly 55% of all churners in the top 20% of scores. This may be interpreted as the Network factors being "responsible" for the 35% out of 55% of churners in the Top 20% of all scores and that correcting these parameters would mitigate at least a part of them[1]. The rest of the churn (the other 20%) is due to other reasons, e.g. a better competitor offer. Having this in mind, it was worthwhile analyzing the parameters that constitute this Purely Network Based model.

[1]In retention campaigns too, one cannot expect 100% acceptance rate

Table 4.4: Univariate performance of predictors (CoC)

Variable	Performance (CoC)
Contract End Date	73.1
Total Duration of Provisioning Errors in the past six months	62.5
Average Ratio of 2G and 3G Data Events in the past three Months	59.2
Count of 2G Data Events in the past three Months	57.5
Sum of Call Drops and Call Setup Failures in the past three months	56.8
Average Voice Call Setup Duration for the past three months	52.4

Due to confidentiality reasons we cannot disclose the exact numbers and weights of the parameters constituting our model. Nevertheless, we can disclose parameters of which our Network model was consisted, ranked by their individual performance (CoC): The Total Duration of Provisioning Errors in the past six months; The Average Ratio of 2G and 3G Data Events in the Past three months; The Count of 2G Data Events in the past three Months; The Sum of Call Drops and Call Setup Failures in the past three months and The Average Voice Call Setup Duration for the past three months. The individual influence (CoC) of each of these parameters onto churn is presented on Table 4.4. Just for comparison, we also show the performance of the best predictor, the contract end date, which has a superior prediction power. However, the purpose of these models was to investigate why customers churn from a network perspective and offer means of alleviating these reasons. In this case, the relationship with contract end date is secondary. When customers get closer to the end of their contract, there is a higher risk of churn. Moreover, customers out of contract for a longer period of time have proven to be loyal, as the other customers have left.

The influence of each of these parameters onto customer experience and therefore churn could be explained and was agreed upon by the company experts. First of all, it is interesting to note that the Sum of Call Drops and Call Setup Failures in the past three months was not a rate, but an absolute count. Namely, it was irrelevant if a customer dropped two calls out of 30 or out of a 100, the two dropped calls drove churn. The parameter Average Voice Call Setup Duration for the past three months implied that customers did not appreciate having to wait a long time to establish a voice call. Provisioning errors are errors where customers have not been enabled to use certain services on the network even though they have subscribed for them (e.g. not being provisioned to use Internet), or did not get the appropriate quality of service (e.g. being provisioned to used Internet at 1 Mbps when subscribed to 3 Mbps). These errors did not occur frequently but were deemed by experts to have a severely negative influence onto satisfaction even if they occurred once during the contract duration; therefore we summed up six months of these errors' history. It is interesting to see the growing influence of mobile Internet services onto churn, especially the

strong preference of customers to use the 3G network, which is by design much faster than the 2G network[2]. The low 2G speed was not deemed satisfactory, it could have been in fact perceived by the customers as not being connected at all. The influence of quality of Internet services onto churn was represented via the Number of 2G Data Events and the Ratio of 3G vs. 2G Data Events.

The added value of these parameters was that they denoted clear guidance for the technology department on which actions to take in order to prevent churn. Determining the exact thresholds of each parameter that led to churn was done by using the discretized variables. As explained in the methodology subsection of this chapter, each variable was separated into bins and each separate bin has gotten a score which is concordant to churn. Next, we were looking for thresholds in these parameters that, once crossed, pointed to higher churn probabilities. For example, let us assume that customers having four or more dropped calls in one month are two times more likely to churn than customers with less than four dropped calls in the same period: This would set the threshold of dropped calls per customer per month to four. This is just a theoretical example, as we are not at liberty to disclose the real figures.

Projects have been developed to maintain and correct these parameters and their respective critical values (increased churn risk thresholds). This also had a profound effect onto the mindset of the department maintaining the network: their focus has shifted from a network centric approach to a customer centric approach in managing the network. We will explain what this means using the example of Voice Call Drops. The network centric approach in managing this key performance indicator would be to just measure a network wide call drop rate and attempt to maintain it above a certain threshold by giving priority to fixing network sites with a large number of dropped calls. The customer centric approach in managing this parameter is to monitor the number of customers experiencing dropped calls and giving highest priority to network sites where most customers experience dropped calls. The customer centric approach allows addressing the problem of a higher number of customers, rather than focusing on network sites where only few customers experience a large number of dropped calls. It has already been implemented and has helped reduce the number of customers experiencing dropped calls in general, which resulted in improved satisfaction in customer surveys (internal to the operator), implicating that churn reduction should follow. Similar approaches were developed to address the other parameters from our model. Last but not least, the technology department in the company has set customer centric targets for managing the network. Using the theoretical example from the paragraph above, this would mean that the department would have set a target that no more than a small percentage of the base (e.g. 1%) should experience four or more dropped calls per month. It is worthwhile mentioning that the amount of customers which was dissatisfied with the operator's network

[2]3G networks could reach throughputs/Internet speeds of 21Mbps, while for 2G the maximum speed was only 64 Kbps

(according to customer surveys) was reducing in parallel with the reduction in the amount of customers experiencing the targeted number of dropped calls and similar customer centric network related targets.

It is possible that the solution applied to a given network site to reduce the number of customers experiencing dropped voice calls may also influence some of the other quality parameters, especially in a case of a 3G network site (e.g. increasing the coverage area or adding extra capacity to a 3G site might reduce both the number of customers experiencing dropped calls and prevent them from falling back to a neighboring 2G cell when using Internet). As an extension of this approach, it can be envisioned that sites where a high number of customers that are already at churn risk experience dropped calls are given priority, but this is subject to legal limitations with regard to data privacy[3].

To summarize, even though our churn model based entirely on network quality parameters had lesser performance compared to a normal campaigning model, it did have many other advantages: it addressed churn in a preventive manner, as it was not necessary to run retention campaigns with it; it provided guidance on which were the critical network parameters that needed to be corrected in order to address churn from a network perspective; and it created a mind-shift in the department managing the network into a customer centric perspective, which already resulted in increased customer satisfaction.

4.5 Limitations and Future Work

The first limitation we would like to address is the lack of coverage data per customer. We were only able to calculate (not measure) the coverage at home for each customer. Loss of coverage for each customer is impossible to measure from the network side. Having adequate coverage information could have improved our model. However, the Ratio between 2G and 3G data events does imply the influence of loss of 3G coverage or insufficient 3G capacity in certain areas onto churn.

Other limitations of this research are of legal nature. Namely, in most European countries stringent Data Privacy or Net Neutrality Laws (will soon) exist. This makes it impossible to look into individual consumption of different types of Internet use (e.g. browsing, streaming, messaging, VoIP etc.), which could provide even better insights into what type of service degradation leads to churn.

Next, as usage patterns change, so do the expectations from the service quality that the network provides. Therefore, in time we expect a change in the influence on churn of the various factors that we discussed which makes the model outdated. This will especially be the case after the introduction of 4G (LTE) networks, which allow much faster Internet speed (throughput). However, these issues can be addressed by remodeling.

[3]It involves storing the sites/locations of particular customers

As future work, we would like to go one step further, and investigate the benefits network experience measured directly on the phone, via a preinstalled app, of course with customers' permission. We believe that this would provide a 360 degrees view of customers' network experience and close the gap created by the data that is difficult to obtain due to technical or legal limitations. Measurements taken directly on the phone are the ultimate determinant of customer's network experience.

4.6 Conclusions

In this chapter we presented an atypical approach to churn management in commercial settings. We succeeded in explaining at least a part of churn via actual measurements of network quality. The main benefits of our approach are the following: First, we managed to build an explanatory churn model by sacrificing only a part of the performance. Second, our churn model was based on features that are extracted from actual network parameters rather than surveys (real network experience vs. perception). Third, this model generated insights on which network parameters are necessary to be corrected in order to reduce churn, which is a new way of churn reduction. The insights generated caused a shift from network centricity towards customer centricity in managing the telecom network. Using this approach, the churn mitigation process is no longer just a retention campaign: the churn efforts are no longer the responsibility of just the CRM teams, Marketing and Customer Service, but also the Technology department, which is responsible for the mobile network is involved. Referring to the research question stated in section 4.1, we have managed to use a different deployment form of a churn model in order explain and prevent churn rather than directly target customers.

Our research was deployed in T-Mobile Netherlands, part of one of the largest European telecom operators. It was used for setting department targets for managing the network and has already contributed to increased customer satisfaction, implicating that churn reduction should follow. In addition, another national operator from the Deutsche Telekom group has used the same approach.

Last but not least, we would like to point out the possibility of applying our research onto domains other than mobile telecom. Obviously, this approach can be mirrored onto fixed telecommunications and potentially into churn in other industries, but also in many other cases where prevention is more important than the cure, like certain medical research.