Applying data mining in telecommunications

Radosavljevik, D.

Chapter 7

Summary and Conclusion

In this thesis we researched various applications of data mining in telecommunications. We had a unique opportunity to do research in a real world setting, using commercial data sets. We worked on broad range of business problems in mobile telecommunications. Our research stretched between the areas of marketing, mobile network technology and finance. We addressed both segments in mobile telecommunications, prepaid and postpaid customers. In chapters 2 and 3 we focused on prepaid customer churn with a target to reduce it by identifying prospective churners. In chapter 4 we built a bridge between a churn model, which is a typically marketing problem, and the mobile network domain in order to prevent it. In chapter 5 we created a simulation platform for mobile network load and in chapter 6 we created a new approach to service revenue forecasting. In this chapter we will summarize the key content and lessons learned from the previous chapters in more detail. We will also address the research questions per chapter and the general research question of this thesis.

In chapter 2, we presented how performance of prepaid churn models changes when varying conditions in three different dimensions: data- by adding CEM (Customer Experience Management) parameters; population sample- by limiting the inactivity period at the time of recording to 15, 30 and zero days, respectively; and outcome definition- by introducing a so-called grace period of 15 days after the time of recording, in which customers must make an activity in order to be classified as churners.

From a theoretical perspective, we created a CEM (Customer Experience Management) Framework for Mobile Telecommunications. Unfortunately, adding the CEM parameters into the models did not add substantial value in terms of model performance under any of the experimental conditions. Similarly, switching the population sample on the period of inactivity at the time of recording between 15 and 30 days did not influence model performance, only the sample size and churn rate. When we changed the population sample by disallowing inactivity at the time of recording, apart from the change in sample size and churn rate there was also a drop in performance and stability of the models. However, this drop in performance was not nearly as high as the one that occurred when changing the outcome definition by setting a grace period, thus making the behavior to be predicted more complex. This change obviously influenced the churn rate as well. Nevertheless, the latter two approaches provided more time for retention. Therefore, looking at the research question posed in section 2.1, which one of the three variations in the experimental setup has the highest influence on prepaid churn modeling, the answer is clearly that changing the outcome definition had a much higher influence on the performance of the prepaid churn models than adding the CEM parameters or changing the characteristics of the sample based on inactivity at the time of recording.

Throughout chapter 3 we have investigated the extent to which social network information can be used to predict telecom churn, and how this information could potentially improve the predictive performance of conventional churn prediction

methods. We have assessed the performance of models constructed using classical tabular data mining, social network mining and hybrid models combining both techniques. The first hybrid model was built by extending the traditional tabular churn predictors with social network variables extracted from the social graph. The second hybrid model was obtained by incorporating the results of a traditional tabular churn model into the social propagation graph, using them as initial energies of the non-churner nodes.

The performance of our models was verified using a large data set of 700 million call data records. Our initial observation showed that the churn probability was positively aligned with the number of churned neighbors. The regular tabular churn models constructed exclusively using social network information and the traditional social network models scored the least. This indicates that social network information alone was not sufficient to predict churn. Overall, the traditional tabular churn models had the best predictive performance. The added value of the social network variables to the tabular churn models was rather minimal. Although the second hybrid models were able to outperform the regular propagation models, they still could not beat the performance of the traditional tabular churn model. The contribution of traditional predictors to churn prediction was substantially higher than that of the social network behavior. Moreover, the performance gain of both hybrid models was not substantial enough to justify the computational costs. In a nutshell, the answer to the research question posed in section 3.1 is that social network mining and attributes stemming from a social network graph did not add substantial value in terms of model performance to traditional prepaid churn modeling in T-Mobile Netherlands.

In chapter 4 we presented an atypical approach to churn management in commercial settings. Utilizing parts of the CEM Framework we created in chapter 2, we succeeded in explaining at least a part of the postpaid churn via actual measurements of network quality. The main benefits of our approach were 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 were extracted from actual network parameters rather than surveys (real network experience vs. perception). Third, this model generated insights on which network parameters were necessary to be corrected in order to reduce churn, which is a new way of churn reduction. This model was built to explain churn and prevent customers from wanting to churn, rather than identifying prospective churners. The generated insights caused a shift from network centricity towards customer centricity in managing the telecom network, meaning focusing on sites where a large number of customers experience network problems rather than on sites where a high number of network problems occur (which could be caused by a malfunctioning of a single phone). Using this approach, the churn mitigation process is no longer just a retention campaign. The churn reduction efforts are no longer the responsibility of just the CRM teams, Marketing and Customer service, but also the Technology department. Managing the

network in a customer centric way is now part of the process and certain customer centric measurements were even set as targets for the Technology department. This resulted into a substantial reduction in the numbers of customers having poor network experience (e.g. the amount of customers experiencing more than 1 dropped call per week has been substantially reduced over two years). Finally, our research has already contributed to reduced dissatisfaction with the network, increased overall customer satisfaction and churn reduction (information proprietary to T-Mobile Netherlands). Therefore, 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.

In chapter 5 we had a goal of using data mining for prediction and simulation of 3G mobile network air interface load. We used a new way to deploy models resulting in a very simple yet effective approach of deploying data mining in commercial surroundings. Unfortunately, data mining is still seen as a black box in many industries, telecom not excluded. Even though some data mining activities are taken, typically in the Marketing/Customer Retention field, there is a myriad of other possibilities in business where data mining can be applied. In our opinion, it is better to start with simple methods, such as linear regression, because it is easier to understand them. Once these simple approaches gain acceptance, and familiarize the industries with data mining, opportunities to apply more advanced techniques will arise. Last but not least, for deployment we used tools that are already familiar to the end-users. This all resulted in a high acceptance of this model and users coming up with their own use cases, which were not originally intended. This model was deployed in multiple national operators of the Deutsche Telekom group. Addressing the research question posed in section 5.1, we have shown how data mining can be used to predict 3G mobile network interface load, and simulate it under different scenarios: we have used relatively simple algorithms to create a large number of predictive models, therefore making possible predicting the load on a cell level. We have used tools known to the end users to deploy these models, allowing them to use different scenarios for input parameters. Acceptance was gained by decoupling the data mining process from the end users, but keeping the transparency that linear regression offers combined with tooling familiar to them.

In chapter 6 we extended this approach onto the field of revenue forecasting. The added value of this model is not only in enabling the business to have much better and much more timely insights in future revenues, assuming that the standard business plans for customer base growth are implemented. This model provides a scenario simulation platform giving the business an opportunity to test the potential measures designed to increase the revenues at much higher pace. Addressing the research question from section 6.1, we have managed to generalize the approach we developed in chapter 5 to the financial domain, and use data mining to predict and simulate service revenues in telecommunications.

Some of the important lessons we learned can be generalized as follows.

In chapter 2 a key lesson is to spend more time defining the problem correctly than picking the right algorithm to solve it or looking for even more data to mine in. This can be a quite interesting finding for many companies who are constantly approached by vendors selling new platforms (more data to mine on) promising to reduce churn across all populations.

Similarly, in chapter 3 we learned that algorithms with high computational cost do not necessarily add predictive power, especially in cases when features traditionally used for the same purpose are already rich.

In chapter 4 we learned that it is worthwhile to try to formulate the problem differently or to look at it from another perspective. The value of an explanatory model can be very high as it can enable one to prevent the problem rather than cure it. Furthermore, the same parameters which did not improve the prepaid churn models in chapter 2, were very useful in designing the explanatory churn model for postpaid customers. This can be seen as a variation of the No Free Lunch theorem related to predictors: Just because certain features did not work on one part of a population (prepaid customers), does not mean they will not have high performance on a different population (postpaid customers). This is not contradictory to the lesson learned from chapter 2: we are just saying that adding more data (a new platform) to mine in does not improve performance on **ALL** problems and populations, but it still may be useful in certain situations.

In chapter 5 we learned that combining simple algorithms in complex deployment forms can reap great benefits. By extending the approach of chapter 5 in chapter 6 we learned that the answer to a company's most important questions can be found by reusing an approach developed for a completely different purpose. In both these chapters we learned that using predictions of inputs for forecasting the output variables created a powerful deployment platform for scenario simulation, resulting in use cases that were not originally foreseen.

In our research we were using a combination of open source software and commercial software. Generally, the commercial software used in some parts of our research can easily be replaced by open source counterparts. However, the automated data preparation (attribute discretization and grouping) as executed in the Predictive Analytics Director software (Pegasystems, 2008), did substantially reduce our workload. The hardware we used was mostly standard of the shelf servers or just a laptop. For deployment we used tools already familiar to the end-users. In general, we did not require additional hardware, software or data from what was already available. Metaphorically speaking, we cooked a meal with what was already in the house. This drives down the cost of applying these solutions, not only in the telecom industry.

From our perspective, the Business Understanding, Data Understanding and Data Preparation stages of the CRISP-DM process were very important. We have learned that in commercial settings the data is very often spread across various sources, so it is essential to unify it. In certain cases, we used existing data sources

and simply imported the necessary data, while in other we created a completely new data repository, which was also useful for operational purposes. Quite often, very rich features already existed it the data, but a so called data dictionary (what does every variable measure) did not. Therefore, domain expert help was crucial for both data preparation and understanding, and of course business understanding. These three stages in an industry setting represent a large part of the overall effort. Having these properly executed results in a rich data set where simple algorithms can perform very well. With regard to data preparation, apart from variable discretization, dealing with missing and extreme values, our choices here were ranging from how much history is relevant to the problem, to generating new variables via different levels of raw data aggregations (hourly, weekly, monthly) and taking into account ratios of these aggregates in order to capture changes in behavior over time.

The deployment step was also very important. This is especially visible in chapters 4, 5 and 6. In chapter 4, the result was not a classic churn model, but rather a set of guidelines for domain experts on what to improve. In chapters 5 and 6, the deployment was actually the key part of the process. Even though the modeling part resulted in a large number of models in a very short time, the method of deploying, which was to first generate new values for the inputs and then use them for forecasting is the part that brought the most of the value. The business now has opportunities to run micro level scenario simulations for two very important business processes. Furthermore, presenting the results using tools familiar to the users also helped the acceptance of the whole process. An overall lesson from a project management perspective was to consult the domain experts who are also the end users of the solutions every step of the way. This is very useful for both setting up the projects at the beginning (at the business understanding, data understanding and preparation stage), as well as for the end- user acceptance. In chapter 5, the users themselves were coming up with new scenarios (use cases)[1]. Last but not least, in business settings the evaluation of models normally does not stop by measuring performance on a test set (pre-labeled data): models are deployed in practice and the performance is benchmarked against actual measured values.

One interpretation of Pareto's rule (Pareto, 1964), especially popular in business, is that 80 percent of the result can be accomplished with 20 percent of the effort (Koch, 2011). One possible translation of this rule to data mining could be that 80 percent of the possible performance improvement can be accomplished with 20 percent of the effort (or time). The timeliness of the solution is often more important than ultimate accuracy. For example, while we are designing the perfect churn model a lot of customers can already be gone. In business, a preferred way is to deploy a solution that is good enough (performs better than the baseline) and improve it later. In some cases, the execution or scoring time of the model is also important: e.g. personalizing a website based on a predictive model: the model cannot cause a high delay in the time necessary to load the page, otherwise the customer might not be patient enough

[1]The approach from Chapter 6 is still under acceptance at the operator

to wait. Applying simple and fast algorithms on inexpensive hardware using open source software can help many organizations that struggle with budgets find at least a temporary solution, which is better than no solution at all (e.g. early detection of diseases).

Looking at the problems we were addressing in this thesis in chronological order, one may conclude that we were progressively addressing problems with higher business importance in each consecutive step. In chapters 2 and 3 we were addressing prepaid churn, which is an important business problem, but in large operators, such as the one where we were conducting our research, prepaid revenues are not nearly as substantial as the revenues from the postpaid segment. The good performance of our models (any of our models were better than the model deployed at the time- not stated in the chapter, as it was out of scope of the research) contributed to getting the task of explaining postpaid churn. After the successful delivery of this task, we were assigned to forecasting network load, which has a huge impact on the budgeting process of a telecom operator, as a large part of the budget is dedicated to network improvements. Last but not least, we got the task to forecast the revenues generated by the postpaid segment of the operator, which is one of the most important financial tasks in a company. One can see this as a journey to accepting data mining in business. Initially, we started with a problem that is important, but not on the top of the list. Solving each consecutive problem was gaining trust, so data mining gained acceptance as a solution to even most important business problems.

To summarize, we believe we have successfully answered the overall research question of this thesis, which was how does one successfully apply data mining in telecommunications? Our approach to this was to focus at the stages of CRISP-DM less covered in literature: business understanding, data understanding and preparation, and in particular deployment. We have used relatively simple algorithms, which performed well on these large datasets and intuitive performance measurements that were easy to explain to the business. We used hardware, software and data which were already available, or added open source software to keep the costs low. We created innovative deployment mechanisms using tools familiar to the end users and involved the users early on in the process. This all has led to the business accepting data mining as a solution to a much broader range of problems than before.

During this research we have paid due attention to the legal and privacy related aspects. European Data Privacy rules are very rigorous about what can and cannot be done with customer data. It is worthwhile mentioning that in chapter 2 and chapter 3 we were working on data from prepaid customers, which do not provide their private information (name, address etc.) to the operator. In chapters 3 and 4, the churn models generated were not used for campaigning. In chapter 3, the intention was to only analyze whether social ties add value to predicting churn. There was no added value, which was an additional reason to not deploy the model. In chapter 4, from the research setup onwards, we were looking for an explanation of churn, not another campaigning model. The results of the model were used to improve

customer experience. For the revenue forecast model in chapter 6, we only used the data that the operator has to store, as mandated by Law, in order to be able to reproduce customer invoices, if necessary. Furthermore, the data used for research in all chapters has been anonymized, except for chapter 5, where we were working on data from network cells, which does not contain any customer identifiers, as this data is already aggregated on a cell level. Unfortunately, in recent times data mining and machine learning techniques have a damaged reputation with relation to privacy (e.g. the case of Edward Snowden, PRISM and NSA; or Google's unification of user profiles for all services; or most recently, Facebook merging WhatsApp and Facebook profile information, for which they were fined €110 million by the EU). We have shown methods to gain valuable insights from customer data which are not in breach of any privacy rights or regulations, neither legally nor ethically.

From a generalization perspective, this thesis is applicable to many industries. Clearly, in almost any industry losing customers to competition (churn), managing the key resource and forecasting revenues are very important problems. Finding inexpensive means to address these issues is beneficial to many companies or governments. Using simple algorithms that can easily be explained and deployment methods preferred by end users helps acceptance. This is how we see data mining being applied and accepted in many fields outside telecommunications, helping more organizations become data driven. From a research perspective, we hope that we have shown that there are many other interesting problems to solve beyond building a better performing predictive model. Hopefully, by applying data mining in telecommunications, we will raise academic interest in finding better and more efficient ways of disseminating machine learning research.