A Case of Churn Prediction in Telecommunications Industry

Brmež, Simon and Žnidaršič, Martin

Abstract: ***Churn prediction is the practice of assigning a probability to the event of a customer ending his contract with a service provider. Traditional data mining approaches to churn prediction in telecommunications industry are based on detecting patterns from customer contractual information, traffic related data, bills and payments, CRM data and customer service logs. The study presented in this paper has employed various machine learning approaches and assessed their performances using the data of a European mobile operator. The feature importance rankings which were used for feature selection yielded also some initial guidelines for acting on churn prevention in practice.***

Index Terms: ***churn, data mining, feature selection, optimization, stacking, telecommunications***

1. Introduction

HISTORICALLY telecommunications industry was product oriented and with abundance of income there was no serious need for customer oriented business environment. Todays' telecommunications companies need to find ways to capture and enhance market shares while reducing costs.

It has been proven, that acquiring new customers costs five to six times more than retaining existing ones [14]. Crucial enabler of targeted retention activities is the ability to perform computerized identification of customers with high propensity to end their relationship with the company, or customer churn prediction. Customer churn is usually a rare event in service industries, but of great interest and great value.Customer churn prediction is the practice of assigning a churn probability to each customer in the company database, according to a predicted relationship between that customer's historical information and its future churning behavior. Practically, the probability to end the relationship with the company is then used to rank the customers from most to least likely to churn, and there can be activities focused on customers with the highest propensity to churn, such as marketing retention campaigns [3].

Traditional data mining approaches to churn detection in telecommunications industry are based on detecting patterns from model input variables derived from customer contractual information, traffic related data, or call detail records, historical records from bills and payments, customer demographics obtained from internal customer relationship management (CRM) systems and customer service logs [15].

This paper presents an exploratory study aimed at churn prediction with machine learning approaches in the context of mobile telecommunications in Slovenia. We experimentally evaluated a selection of algorithms with respect to their churn prediction performance. Some parts of our study, such as the feature rankings, provided also initial operational guidelines for acting towards churn prevention. This work represents initial steps in an effort to analyze the usage and general cus-

Manuscript received May 2019.

The authors acknowledge the financial support from the Slovenian Research Agency for research core funding for the programme Knowledge Technologies (No. P2-0103)

S. Brmež (simon.brmez@gmail.com) is with the Jožef Stefan International Postgraduate School, Ljubljana, Slovenia. M. Žnidaršič (martin.znidarsic@ijs.si) is with the Jožef Stefan Institute, Ljubljana, Slovenia and with the Jožef Stefan International Postgraduate School, Ljubljana, Slovenia

tomer data, as well as the contents of customer feedback information from various communication chanells (calls, e-mails) and contextual data (time, season, etc.) in order to better understand the churn phenomenon in the specific telecommunication industry and the factors that influence it. The knowledge, algorithms and models gained by these studies are to be applied in practice with an aim to better understand the needs of the customers and reduce the churn and its negative effects.

Slovenia is one of the smallest markets in the European Union and for its context and demography the studies of churn in telecommunications are rare (see e.g. [4]) and to the best of the authors knowledge there is no published research on predictive performance of machine learning approaches in such a context.

2. Related Work

Hadden et al. [6] define five stages of a churn management framework:

1. Identification of the most appropriate data,
2. Data semantics,
3. Feature selection,
4. Development of a predictive model,
5. Validation of results.

Identification of the most appropriate data is the initial step in developing a customer churn management framework as defined by Hadden et al. [6]. As different combinations of data hold different analytical powers, it is necessary to identify the data that best suits the analysis being performed.

In case of prepaid customers in the cellular telecommunication industry, as described by Owczarczuk [12], only usage data such as call detail records and service options activations or deactivations can be used as data for predictive modeling. In most countries prepaid customers are anonymous and there is no demographic data available. On the other hand, when dealing with contract based postpaid customers, demographic data can be successfully utilized alongside usage data for predictive modeling as shown by Kisioglu and Topcu [10].

Data semantics is the process of understanding the context of the data in a database, as defined by Hadden et al. [6] and further described as objects, relationships amongst objects, and properties of objects. Coussement et al. [3] follow the established Cross-Industry Standard Process for Data Mining (CRISP-DM) and break it down into six distinct stages: business understanding, data understanding, data preprocessing, modeling, evaluation, and deployment.

Feature selection is the critical process of identifying the variables which are the most relevant for prediction. It is an important stage because it helps with both data cleansing and data reduction, by including the important features and excluding the redundant, noisy and less informative ones. A widely used method for feature selection is Relief by Kira and Rendell. A variant of this method with several extensions proposed by Kononenko [11] is called ReliefF. Enormous size, high dimensionality and imbalanced nature of telecommunication datasets are main hurdles in attaining the desired performance for churn prediction. Idris et al. [8] investigated the significance of a Particle Swarm Optimization (PSO) based undersampling method to handle the imbalance data distribution in collaboration with different feature reduction methods: Principle Component Analysis (PCA), Fisher's ratio, F-score and Minimum Redundancy and Maximum Relevance (mRMR). Random Forest (RF) and K-Nearest Neighbour (KNN) classifiers were employed to evaluate the performance on optimally sampled and reduced features dataset. They observed through simulations that their proposed approach based on PSO, mRMR, and RF performed satisfactory for churn prediction problem. Class imbalance problem in customer churn prediction was researched by Burez and Van den Poel [1] and they showed that undersampling can lead to improved prediction accuracy. However they also confirmed that there is no general answer as to which class distribution will perform best, and that the answer is surely method and case dependent.

Development of a predictive model is the next stage in churn management framework and defined by Hadden et al. [6] as one that takes patterns that have been discovered in the database, and predicts the future values. Keramati et al. [9] employed data mining classification techniques including Decision Tree (DT), Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) so as to compare their performances as predictive models using the data of an Iranian mobile company. Evaluation of performances of each particular algorithm was done using F-measure with average values ranging from 0.749 for KNN up to 0.8622 for ANN, DT with 0.8382 and SVM with 0.8286. Superior classification accuracy was achieved with hybrid approach including all four techniques with considerably higher than 95% accuracy for both precision and recall measures. Huang and Ketchadi [7] also proved, that hybrid approach can yield better classification results as opposed to using only single classification technique. Coussement and De Bock [2] similarly demonstrated the beneficial effect of ensemble learning method. In their research single CART decision trees were benchmarked to their ensemble counterparts, random forests. On the other hand, Vafeiadis et al. [13] studied several methods (ANN, SVM, DT, Naive Bayes and Logistic Regression) including wide ranges of parameters for each method. Their results demonstrate clear superiority of the boosted versions of the models against the plain (non-boosted) versions. The best overall classifier was the SVM-POLY using AdaBoost with accuracy of almost 97% and F-measure over 84%. Droftina et al. [4] propose interesting approach to churn prediction using diffusion model based on sociometric clique and social status theory and its superior performance was demonstrated on a real dataset of users obtained from the largest Slovenian mobile service provider.

Validation of results is the last step in churn management framework described by Hadden et al. [6]. Variations of cross-validation are the most popular. Idris et al. [8] used 10-fold cross-validation in their experiments. Huang et al. [7] used 10-fold cross-validation and also compared the performances of algorithms using ROC curves. Coussement et al. [3] used splits on research data of 50% for training set, a selection set (20%), and a validation set (30%). Lift curves are also very informative methods of visualizing the results and comparing performances as used by Owczarczuk [12] in his research.

3. Experimental setup and results

The aim of the experiments conducted in the scope of this study is to provide initial insights into the characteristics of the analysed data and the applicability of the assessed machine learning methods. In the following we describe the data, the feature selection approach and the machine learning methods that were used in our experimental work.

3.1 Data

Experimental setting included data from a European mobile telecommunications operator for residential customer segment. 3 months of usage data for September, October and November 2016 was aggregated to subscription level yielding 505.301 subscriptions belonging to 405.036 distinct customers. 6.516 subscriptions were labeled as churned in the month of December 2016 from the CRM database. The criterion for being a churner was the termination of subscription's contract in month of December 2016. These 6.516 subscriptions represented the target class for prediction and belonged to 5.953 distinct customers.

All datasets were anonymized in accordance with general European personal data protection laws and prepared in a way that not one single customer could be identified in reverse.

In total, usage data initially contained 86 variables, 1 being the target variable:

- 4 variables as identifiers (subscription, customer, period and source system);
- 4 variables for various billed amounts;
- 4 variables describing additional binding contracts;
- 13 variables for various contract data from CRM system;

- 17 variables for technical properties of the mostly used handset in particular period;
- 4 variables for counts and spent amounts for messaging services;
- 9 variables for usage of mobile packet data service;
- 9 variables describing the magnitude of social circle, such as count of distinct numbers called in competitor's network;
- 2 variables describing commercial tariff model;
- 19 variables describing durations and amounts of different types of voice traffic.

3.2 Feature Selection

The feature selection experiments were done on an independent traffic and CRM data from August 2016 and predicting churn event in September 2016. As proposed by Keramati el al. [9], we employed SVM (linear) as the baseline classifier for the feature selection process. The feature ranking algorithms used were: (I) ReliefF from an implementation in R and (II) the forests of trees feature importance approach from the scikit-learn library. ReliefF determined that 64 features have a positive score and thus have a positive influence on the target variable. The forests of trees feature importance approach was somewhat less restrictive with only four features evaluated with a zero score. The two rankings of features proved to be slightly diverse, but with a lot of overlap among the most highly ranked features. The feature selection for our purposes was conducted according to the results of ReliefF by selecting all the features with a positive score. Top 10 features accoring to ReliefF (all have a score above 0.15) are listed below, with the features that are also amongst the top 10 features identified by forests of trees marked with an asterisk (*):

- commercial tariff bundle (*);
- reason for contract subscription (*);
- relative date of starting the contract (*);
- total number of open bindings;
- relative starting date of the last binding contract (*);
- relative end date of the last binding contract (*);
- handset type used (*);
- number of commercial tariff bundle changes;
- total number of binding contracts;
- support for 3G/4G on handset.

3.3 Experimental Process and Results

The traffic and CRM dataset that was used in the experimets consisted of 6.516 subscriptions labeled as churned with additional 10.000 random subscriptions from majority class as undersample, thus removing the majority class examples as one of the popular methods to tackle the class imbalance problem [1]. In the experimental assessments, a stratified 3-fold cross validation was used to separate the data into the training and testing parts.

We conducted four sets of experiments to assess the baseline results and performance of various machine learning algorithms, the effects of feature selection and parameter optimization, the potential benefits of ensamble approaches and the extension of the analyzed period before predicting churn events.

Initial experiments

The initial set of experiments was done on traffic and CRM data from November 2016 and churn events from December 2016. The goal was to establish a baseline using traditional approach to churn prediction in telecommunications as was done by Owczarczuk [12] and Keramati et al. [9]. Experiments were done with dataset containing all the features and default settings for each algorithm. Compared learners were: linear SVM (SVM linear), SVM with a radial basis function kernel (SVM rbf), random forest (RF), K-nearest neighbors (KNN) and logistic regression (LR). The results of these experiments are shown in Table 1.

Table 1: Results of the initial run of experiments.

	SVM linear	SVM rbf	RF	KNN	LR
Classification Accuracy	0.662	0.655	0.698	0.683	0.672
Recall	0.650	0.730	0.390	0.450	0.400
Precision	0.500	0.500	0.600	0.540	0.533

Parameter optimization

In the second run of experiments, the feature selection was employed (see description in Section 3.2) as well as parameter optimization, which was done using the scikit-learn's grid search parameter tuning method. Only the features indicated as relevant by ReliefF were used in this and in subsequent experiments. Results of this run of experiments are presented in Table 2. The SVM and the logistic regression learners are not sensitive with regards to redundant features, so their results are not positively affected, while the results of random forest and K-NN improved.

Stacking approach

The third set of experiments was dedicated to the assessment of an ensamble approach of stacking of the classifiers. We assessed three combinations: SVM rbf + RF, SVM rbf + RF + KNN and RF + KNN. In all cases, the logistic regression was used as the meta classifier. Results of this set of experiments for each combination of classifiers are shown in Table 3. The best result, which outperforms the results of any individual classifier, was achieved by using the stacking approach with random forest and the SVM with an RBF kernel as base classifiers. This confirms the findings of Keramati et al. [9] and Huang and Kechadi [7] that ensamble approaches outperform the usage of solely base classifiers.

Extended observation period

After discussion of experimental results and the characteristics of the studied churn phenomena with domain experts we came to the conclusion that we should consider and analyze more months of data prior to churn events. A suggestion was also made that an emphasis should be put on the recall measure, so we introduced the F-2 measure as the overall indicator of predictive preformance. Stacking classifier approach with traffic and CRM dataset resulted in F-2 score of 0.6229. Table 4 shows the experimental results for the cases of using 1, 2 and 3 months of data prior to the churn event in terms of the F-2 measure. Extension of the observation period to 2 months seems to be beneficial, while additional extension beyond that does not improve the results.

Table 4: Experimental results when considering longer periods of analysis.

Months	F-2
1	0.623
2	0.665
3	0.662

4. Conclusion and Future Work

Our research and subsequent discussions with domain experts and practitioners have affirmed the practical usefulness of data-based churn prediction in our specific setting. We also confirmed the results of related works, for example that ensemble approach yields better classification results opposed to using single classifiers in such a context. However, we do not yet reach the performances as reported in related works, either due to the limited nature of our study or due to demographic specifics. We intend to extend the scope of data used and further improve the methodology to try to improve on that.

Our results indicate that (perhaps due to tighter telecommunuications market regulation) features derived from CRM data, such as contract bidings data, play more important role in churn detection than behaviour features derived from traffic data and that special attention should be payed in the period nearing the contract binding ending dates. This period should be longer than one month, as classfica-

Table 2: Results after feature selection and parameter optimization.

	SVM linear	SVM rbf	RFF	KNN	LR
Classification Accuracy	0.662	0.652	0.724	0.703	0.674
Recall	0.640	0.620	0.480	0.390	0.400
Precision	0.500	0.490	0.620	0.600	0.530

Table 3: Experimental results of three combinations of stacking.

	SVM rbf + RF	SVM rbf + RF + KNN	RF + KNN
Classification Accuracy	0.731	0.720	0.710
Recall	0.500	0.330	0.370
Precision	0.640	0.680	0.650

tion performance improved when we used periods of 2 and 3 months before a churn event.

Possible extensions to churn prediction could include: social network analysis as proposed by [4], text analysis of customer care center e-mail messages and sentiment analysis as proposed by Duric [5] on various channels where customers leave comments.

References

[1] Jonathan Burez and Dirk Van den Poel. Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36(3):4626–4636, 2009.

[2] Kristof Coussement and Koen W De Bock. Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning. *Journal of Business Research*, 66(9):1629–1636, 2013.

[3] Kristof Coussement, Stefan Lessmann, and Geert Verstraeten. A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decision Support Systems*, 95:27–36, 2017.

[4] Uroš Droftina, Mitja Štular, and Andrej Košir. A diffusion model for churn prediction based on sociometric theory. *Advances in Data Analysis and Classification*, 9(3):341–365, 2015.

[5] Adnan Duric and Fei Song. Feature selection for sentiment analysis based on content and syntax models. *Decision support systems*, 53(4):704–711, 2012.

[6] John Hadden, Ashutosh Tiwari, Rajkumar Roy, and Dymitr Ruta. Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, 34(10):2902–2917, 2007.

[7] Ying Huang and Tahar Kechadi. An effective hybrid learning system for telecommunication churn prediction. *Expert Systems with Applications*, 40(14):5635–5647, 2013.

[8] Adnan Idris, Muhammad Rizwan, and Asifullah Khan. Churn prediction in telecom using random forest and pso based data balancing in combination with various feature selection strategies. *Computers & Electrical Engineering*, 38(6):1808–1819, 2012.

[9] Abbas Keramati, Ruholla Jafari-Marandi, Mohammed Aliannejadi, Iman Ahmadian, Mahdieh Mozaffari, and Uldoz Abbasi. Improved churn prediction in telecommunication industry using data mining techniques. *Applied Soft Computing*, 24:994–1012, 2014.

[10] Pınar Kisioglu and Y Ilker Topcu. Applying bayesian belief network approach to customer churn analysis: A case study on the telecom industry of turkey. *Expert Systems with Applications*, 38(6):7151–7157, 2011.

[11] Igor Kononenko. Estimating attributes: analysis and extensions of relief. In *European conference on machine learning*, pages 171–182. Springer, 1994.

[12] Marcin Owczarczuk. Churn models for prepaid customers in the cellular telecommunication industry using large data marts. *Expert Systems with Applications*, 37(6):4710–4712, 2010.

[13] Thanasis Vafeiadis, Konstantinos I Diamantaras, George Sarigiannidis, and K Ch Chatzisavvas. A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55:1–9, 2015.

[14] Wouter Verbeke, Karel Dejaeger, David Martens, Joon Hur, and Bart Baesens. New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, 218(1):211–229, 2012.

[15] Chih-Ping Wei and I-Tang Chiu. Turning telecommunications call details to churn prediction: a data mining approach. *Expert systems with applications*, 23(2):103–112, 2002.

Simon Brmež received his masters degree in Decision Support Systems from the University of Maribor in 2011, He has 20 years of expirience building enterprise scale business intelligence systems in telecommunications industry. His research is focused on data mining, text analytics and sentiment analysis.

Martin Žnidaršič received his doctoral degree in Computer and Information Science from the University of Ljubljana in 2007. He is currently employed at the Department of Knowledge Technologies at Jožef Stefan Insltitute, Ljubljana, Slovenia. His research is focused on decision support and data mining, in particular on human assessment modelling and sentiment analysis. Applications of his work are conducted mostly in domains of ecology and finance.