Churn Prediction

Vladislav Lazarov
Technische Universität München
vladislav.lazarov@in.tum.de

Marius Capota
Technische Universität München
mariuscapota@yahoo.com

ABSTRACT
The rapid growth of the market in every sector is leading to a bigger subscriber base for service providers. More competitors, new and innovative business models and better services are increasing the cost of customer acquisition. In this environment service providers have realized the importance of the retention of existing customers. Therefore, providers are forced to put more efforts for prediction and prevention of churn. This paper aims to present commonly used data mining techniques for the identification of churn. Based on historical data these methods try to find patterns which can point out possible churners. Well-known techniques used for this are Regression analysis, Decision Trees, Neural Networks and Rule based learning. In section 1 we give a short introduction describing the current state of the market, then in section 2 a definition of customer churn, its' types and the imporance of identification of churners is being discussed. Section 3 reviews different techniques used, pointing out advantages and disadvantages. Finally, current state of research and new emerging algorithms are being presented.

Categories and Subject Descriptors
H.2.8 [**Database Management**]: Database Applications—*Data mining*

General Terms
Algorithms, Human factors, Management, Theory

Keywords
Churn, Churn prediction, Decision trees, Neural networks, Regression

1. INTRODUCTION
In a world of ever growing competition on the market, companies have become aware that they should put much effort not only trying to convince customers to sign contracts, but also to retain existing clients. Van Den Poel and Larivière [2] have shown that in the current setting where people are given a huge choice of offers and different service providers to decide upon, winning new customers is a costly and hard process. Therefore, putting more effort in keeping churn low has become essential for service-oriented companies.

Van Den Poel and Larivière [2] summarize the economic value of customer retention:

- lowering the need to seek new and potentially risky customers, which allows focusing on the demands of existing customers;
- long-term customers tend to buy more;
- positive word of mouth from satisfied customers is a good way for new customers' acquisition;
- long-term customers are less costly to serve, because of a larger database of their demands;
- long-term customers are less sensitive to competitors' marketing activities;
- losing customers results in less sales and an increased need to attract new customers, which is five to six times more expensive than the money spent for retention of existing customers;
- people tend to share more often negative than positive service experience with friends, resulting in negative image of the company among possible future customers.

Customer Relationship Management (CRM) tools have been developed and applied in order improve customer acquistion and retention, increase of profitability and to support important analytical tasks such as predictive modelling and classification. Typically, CRM applications hold a huge set of information regarding each individual customer. This information is gained from customers' activity at the company, data entered by the customer in the process of registration, calls to support hotlines, etc. Proper analysis of this data can bring remarkable results for marketing purposes, but also for identifying customers which are likely to cancel their contract.

Typically, database entries are scored using a statistical model defined over various attributes, which characterize the customers. These attributes are often called predictor variables.

Higher scores reveal greater possibility of churning. Models are being built using statistical techniques like regression analysis, classification trees and neural networks.

This paper tries to investigate common applications and methodologies for churn prediction, pointing out assets and drawbacks of each of them.

2. CUSTOMER CHURN

'Churn' is a word derived from *change* and *turn*. It means the discontinuation of a contract.

There are three types of churn:

- **active / deliberate** - the customer decides to quit his contract and to switch to another provider. Reasons for this may include: unsatisfaction with the quality of service (e.g. not fulfilling service level agreements), too high costs, not competitive price plans, no rewards for customer loyalty, no understanding of the service scheme, bad support, no information about reasons and predicted resolution time for service problems, no continuity or fault resolution, privacy concerns, etc.
- **rotational / incidental** - the customer quits contract without the aim of switching to a competitor. Reasons for this are changes in the circumstances that prevent the customer from further requiring the service, e.g. financial problems, leading to impossibility of payment; or change of the geographical location of the customer to a place where the company is not present or the service is unavailable.
- **passive / non-voluntary** - the company discontinues the contract itself.

Voluntary churn (active, rotational) is hard to predict. And while incidental churn only explains a small fraction of overal churn it is particularly interesting to predict and react taking appropriate action to prevent deliberate churn. In order to prevent customers' voluntary contract discontinuation, however, the company needs to know **who** are the possible churners with a low probability of error in the prediction and **why** this specific customer has decided to leave the company for the benefit of a competitor.

Furthermore, churning can be divided also in three other groups:

- **total** - the agreement is officially cancelled;
- **hidden** - the contract is not cancelled, but the customer is not actively using the service since a long period of time;
- **partial** - the agreement is not cancelled, but the customer is not using the services to a full extent and is using only parts of it, and is instead using constantly a service of a competitor.

Depending on the company, the contract type and the business model that is being applied hidden or partial churning can lead to considerable money loss (e.g. in telecommunications: the customer only pays the monthly subscription fee, but does not place a single call) and also needs to be identified and action should be taken in order not to loose completely the customer.

Moreover, it is important to classify which of the possible churners are of further interest for the company, e.g. which customers are likely to generate more profit (these are typically customers who generated substantial revenues and then found a better offer with a good loyalty programme at a competitor), and which customers are not interesting, because, for instance, they are identified as risky. Then the company marketing department can consider direct marketing strategies in order to retain important customers.

Although churn as a whole is an unavoidable phenomenon, it can be managed and the potential losses to the business can be minimized. The timely detection of possible churners, together with effective retention efforts support this goal.

3. CHURN PREDICTION METHODOLOGY

For finding answers to the questions who and why is likely to churn a classification of the customers is needed. Churn prediction deals, therefore, with the identification of customers likely to churn in the near future. The basis for this is historical data, containing information about past churners. A comparison is made between these churners and existing customers. As likely churners are identified customers for which the classification suggests similarity to prior churners. Mitchell [7] summarizes different terms used for construction of a classification procedure as pattern recognition, discrimination or supervised learning.

3.1 Data Set

Service providers can easily acquire huge volumes of data. Van den Poel and Larivière [2] present four sets of data variables: customer behaviour, customer perceptions, customer demographics and macroenvironment variables.

- **Customer behaviour** identifies which parts of the service a customer is using and how often is he using them. Interesting are product-specific ownership (which product/service is owned/on loan), total product ownership (number of products owned/on loan), interpurchase time (time between the purchase of two different articles). In telecommunications, for example, the provider can track number and length of calls, period between calls, the usage of the network for data exchange, etc.;
- **Customer perceptions** are defined as the way a customer apprehends the service. They can be measured with customer surveys and include data like overall satisfaction, quality of service, problem experience, satisfaction with problem handling, pricing, locational convenience, image/reputation of the company, customer perception of dependency to the vendor, etc.;
- **Customer demographics** are some of the most used variables for churn prediction. They include age, gender, level of education, social status, geographical data, etc.;

- **Macroenvironment variables** identify changes in the world, different experiences of customers, which can affect the way they use a service. For example, in the telecommunication industry people who have survived a natural disaster and could rely on their mobile phones during it are more likely to continue using the service.

The size of gathered data is usually very large, which results in high dimensionality, making the analyze a complex and challenging task. Therefore, before beginning to use a churn prediction method a data reduction technique is used, deciding with application domain knowledge which attributes can be of use and which can be ignored. Missing values should also be regarded - on attribute level these can be ignored if they are with low significance, whereas on record level they have to be replaced with a reasonable estimate, for example using interpolation. Providing a good estimate for this missing values is an important issue for proper churn prediction.

3.2 Typical Situation

The typical approach to the problem of churn prediction is using a sufficiently large data set, that contains churning and non-churning customers. This set is being analyzed to construct a classifier. The work of a classifier is to decide, given a customer data set, if churn is more or less likely. Such classifiers are constructed using, for instance, neural networks, bayesian statistics or decision trees constructed with the he heuristics like CART or C4.5 [8].

	Actual Churners	Actual Non Churners
Predicted Churners	True Positive	False Positive
Predicted Non Churners	False Negative	True Negative

Table 1: Churn prediction categories

Quality of the output is then measured in terms of sensitivity, specifity and accuracy. Table 1 shows the categorization of churn prediction. The *sensitivity* of a classifier is the number of data sets for which correct predictions have been made (true positives, in our case: churners predicted as churners) divided by the total number of members (true positives + false positives). The *specifity* is the number of data sets that were correctly predicted to not be members of the class (true negatives, non churners predicted as non churners) divided by the number of all members that do not belong to the class. Usually a Receiver Operating Characteristic (ROC) curve is used to display a graphic of sensitivity vs. specifty. *Accuracy* is defined as the percentage of correct predictions.

This quality measures are used to adjust parameters of the classifier until a reasonable quality of prediction is achieved. An accuracy of about 90% is, according to Domingos [3], sufficient for a classifier to predict churning.

3.2.1 Regression Analysis

Regression is considered to be a good technique for identifiying and predicting customer satisfaction. For each of the variables in a regression model the standard error rate is calculated using SPSS. Then the variables with the most significance in respect to linear regressions for churn prediction are obtained and a regression model is constructed.

Since the prediction task in churn prognosis is to identify a customer as a churner or non churner and therefore the prediction attribute is associated with only two values logistic regression techniques are suitable. While linear regression models are useful for prediction of continuous valued attributes, logistic regression models are suitable for binary attributes.

The logistic regression model is simply a non-linear transformation of a liner regression model. The standard representation of logistic regression is referred as logit function.

The estimated probability of churn is estimated with the function

$$Pr[churn] = \frac{1}{1+e^{-T}} \tag{1}$$

where $T = a+BX$. Here a is a constant term, X represents the predictor attributes vector and B is the coefficient vector for the predictor attributes. If T equals 0 the probability is 0,5. This means that it is equi-probable that a customer is a churner and non churner. With T growing large the probability comes closer to 1, so the customer becomes a more probable churner, when T is becoming small the probability of churn is tending to be 0.

3.2.2 Naive Bayes

Naive Bayes is a type of supervised-learning module that contains examples of the input-target mapping the model tries to learn. Such models make predictions about new data based on the examination of previous data. The Naive Bayes algorithm uses the mathematics of Bayes' Theorem to make its predictions.

$$Pr[A|B] = \frac{Pr[B|A]\,Pr[A])}{Pr[B]} \tag{2}$$

Bayes' Theorem (2) states that the probability of a particular predicted event, given the evidence in this instance, is computed from three other numbers: the probability of that prediction in similar situations in general, ignoring the specific evidence (the so called prior probability) multiplied with the probability of seeing the evidence we have here, given that the particular prediction is correct divided by the probability of that prediction in general.

3.2.3 Decision Trees

Decision trees are the most commonly used tool for predictions and classification of future events. The development of such trees is done in two major steps: building and pruning. During the first phase the data set is partitioned recursevely until most of the records in each partition contain identical value. The second phase then removes some branches which contain noisy data (those with the largest estimated error rate).

CART, a Classification And Regression Tree, is constructed by recursive splits of an instance into subgroups until a specified criteria has been met. The tree grows until the decrease of impurity falls below a user-defined threshold.

Each node in a decision tree is a test condition and the branching is based on the value of the attribute being tested. The tree is representing a collection of multiple rule sets. When evaluating a customer data set the classification is done by traversing through the tree until a leaf node is reached. The label of this leaf node (Churner or Non Churner) is assigned to the customer record under evaluation.

Figure 1 shows a simplified churn prediction decision tree for the telecommunication sector.

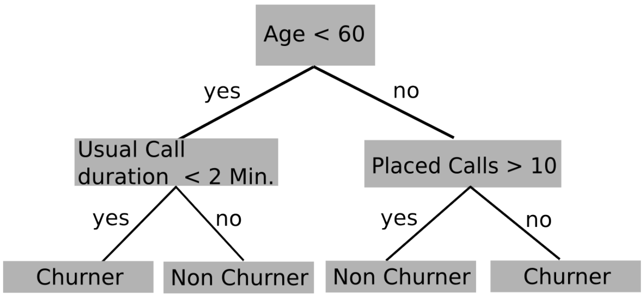

Figure 1: A simplified churn prediction decision tree

Decision trees are often criticised that they are not suitable for capturing complex and non-linear relationshis between the attributes. Nevertheless, research shows [4, 5, 8] that the accuracy of decision trees and training data requirements are high.

3.2.4 Neural Networks

The use of neural networks in churn prediction has a big asset in respect of other methods used, because the likelihood of each classification made can also be determined. Au *et al.* have shown [1] that neural networks outperform decision trees for prediction of churn. They state the biggest disadvantage of neural networks – they do not uncover patterns in an easily understandable form, categorizing them as a 'black box' model

The basic idea behind neural networks is that each attribute is associated with a weight and combinations of weighted attributes participate in the prediction task. During learning the weights are constantly updated, thus correcting the 'effect' which an attribute has. Given a customer data set and the set of predictor variables the neural network tries to calculate a combination of the inputs and to output the probability that the customer is a churner.

Accuracy achieved by neural networks fully outweights the disadvantage that they need a large volume of data set and a lot of time in order to calculate a reasonable weightage for the predictor attributes.

Iwata *et al.* [6] have proved that neural networks are superior in performance as opposed to other models. Similar results where identified by Hadden *et al.* in [4, 5].

3.3 Problems of Churn Prediction

Although conventional churn prediction techniques have the advantage of being simple and robust with respect to defects in the input data, they posses serious limitations to the interpretation of reasons for churn. Therefore, the answer of the question **who** is probably going to churn is easily given, however the bigger problem is to find **why** the customer wants to churn. A proper insight on the reasons for churning is essential in order to design effective retention methods. Therefore, measuring the effectiveness of a prediction model depends also on how well the results can be interpreted for inferring the possible reasons of churn. This is done in order to properly allocate limited time and ressources for retention efforts by choosing more probable churners with the help of successful (in previous efforts for the given reason) techniques.

Neural networks, being best in terms of performance have proven to be best for actual churn prediction. The results can then be further analyzed using linear regression and decision trees for explaining the behavior of churn.

3.3.1 Evolutionary Approach

The necessity of precise interpretation of churn prediction results motivated researchers for the suggestion of new and precise models, which not only offer insights whether a customer is likely to churn or not, but also focus on reasons for churning.

An evolutionary algorithm learning of rules for churn has been proposed by Au *et al.* [1]. The Algorithm is described by the authors as follows: 1) the evolutionary process begins with the generation of an initial set of first-order rules (i.e., rules with one conjunct/condition) using a probabilistic induction technique and based on these rules, rules of higher order (two or more conjuncts) are obtained iteratively; 2) when identifying interesting rules, an objective interestingness measure is used; 3) the fitness of a chromosome is defined in terms of the probability that the attribute values of a record can be correctly determined using the rules it encodes; and 4) the likelihood of predictions (or classifications) made are estimated so that subscribers can be ranked according to their likelihood to churn.

3.3.2 Self-organizing Maps

Ultsch describes [10] how a combination of emergent self organizing maps, U-Matrix methods and knowledge conversion can be used for churn prediction and conversion, creating an effective churn prediction classifier. In this approach groups amongst customers are identified. Group characteristics are being summarized using rules. Research results have shown a correct prediction of 90%. The main idea is that identification of groups and their characteristics can lead to better understanding of the real reasons for churning.

4. WEIGHTING RESULTS WITH CUSTOMER LIFETIME VALUE

Once churners have been identified and reasons for quitting have been found rapid action has to be taken by the marketing department in order to prevent churn properly. Usually the time is not enough to address all likely churners. Therefore, further decision making has to be done to choose the clients that will be contacted. For the customers with highest probability of churning it is predicted how much revenue a service provider is going to get over the period of customers' stay. In this way valuable customers are identified and efforts are made of retaining these customers. An inversely proportional rate of customer lifetime value of an existing customer to the churn probability of that customer can be seen [9], but the customers' decision to stay back is usually coupled with increment of lifetime value of that customer.

Using customer lifetime value in addition to churn prediction can minimize the cost for making a needless retention effort (false positives) and the cost of losing a customer because the model did not predict he is likely to churn (false negatives).

5. SUMMARY

Retention of possibly churning customers' has emerged to be as important for service providers as the acquisition of new customers. High churn rates and substantial revenue loss due to churning have turned correct churn prediction and prevention to a vital business process. Although churn is unavoidable, it can be managed and kept in acceptable level. There are many different ways of churn prediction and new techniques continue to emerge. Good prediction models have to be constantly developed and a combination of the proposed techniques has to be used. Valuable customers have to be identified, thus leading to a combination of churn prediction methods with customer lifetime value techniques.

In this paper we presented typical quality measures of prediction models, together with 6 churn prediction methods: regression analysis, naive bayes, decision trees, neural networks, new evolutionary approaches and self-organizing maps combined with U-Matrix. Furthermore, we pointed out links between churn prediction and customer lifetime value.

6. REFERENCES

[1] W.-H. Au, K. C. C. Chan, and X. Yao. A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Trans. Evolutionary Computation*, 7(6):532–545, 2003.

[2] D. V. den Poel and B. Larivière. Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, 157(1):196–217, 2004.

[3] P. Domingos. The role of occam's razor in knowledge discovery. *Data Min. Knowl. Discov.*, 3(4):409–425, 1999.

[4] J. Hadden, A. Tiwari, R. Roy, and D. Ruta. Churn Prediction: Does Technology Matter. *International Journal of Intelligent Technology*, 1(1):104–110, 2006.

[5] J. Hadden, A. Tiwari, R. Roy, and D. Ruta. Churn Prediction using Complaints Data. *International Journal of Intelligent Technology*, 13:158–163, May 2006.

[6] T. Iwata, K. Saito, and T. Yamada. Recommendation method for extending subscription periods. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 574–579, New York, NY, USA, 2006. ACM.

[7] T. Mitchell. *Machine Learning*. WCB/Mc Graw Hill, Boston, et al., 1997.

[8] J. R. Quinlan. *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

[9] S. Rosset, E. Neumann, U. Eick, N. Vatnik, and Y. Idan. Customer lifetime value modeling and its use for customer retention planning. In *KDD Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 23-26, 2002, Edmonton, Alberta, Canada*, pages 332–340, 2002.

[10] A. Ultsch. Data mining and knowledge discovery with emergent self-organizing feature maps for multivariate time series. In E. Oja and S. Kaski, editors, *Kohonen Maps*, pages 33–45. Elsevier, 1999.