Chapter XIII
New Churn Prediction Strategies in Telecom Industry

Dymitr Ruta
BT Group, Research and Venturing, UK

Christoph Adl
BT Group, Research and Venturing, UK

Detlef Nauck
BT Group, Research and Venturing, UK

ABSTRACT

In the telecom industry, high installation and marketing costs make it six to 10 times more expensive to acquire a new customer than it is to retain an existing one. Prediction and prevention of customer churn is therefore a key priority for industrial research. While all the motives of customer decision to churn are highly uncertain there is a lot of related temporal data generated as a result of customer interaction with the service provider. The major problem with this data is its time discontinuity resulting from the transactional character of events they describe. Moreover, such irregular temporal data sequences are typically a chaotic mixture of different data types, which further hinders its exploitation for any predictive task. Existing churn prediction methods like decision trees typically classify customers into churners and non-churners based on the static data collected in a snapshot of time while completely ignoring the timing of churn and hence the circumstances of this event. In this work, we propose new churn prediction strategies that are suitable for application at different levels of the information content available in customers’ data. Gradually enriching the data information content from the prior churn rate and lifetime expectancy then typical static events data up to decay-weighted data sequences, we propose a set of new churn prediction tools based on: customer lifetime modelling, hidden markov model (HMM) of customer events, and the most powerful k nearest sequence (kNS) algorithm that deliver robust churn predictions at different levels of data availability. Focussing further on kNS we demonstrate how the sequential
INTRODUCTION

Today’s global telecommunication market environment can be characterised by the strong competition among different telcos and a decline in growth rate due to maturity of the market. Furthermore, there is a huge pressure on those companies to make healthy profits and increase their market shares. Most telecom companies are in fact customer-centric service providers and offer to their customers a variety of subscription services. One of the major issues in such environment is customer churn known as a process by which a company loses a customer to a competitor. Recent estimates suggest that churn rates in the telecom industry could be anywhere from 25 percent to 50 percent (Furnas, 2003). Moreover on average it costs around $400 to acquire a new customer, which takes years to recoup (Furnas, 2003). These huge acquisition costs are estimated to be five to eight times higher than it is to retain the existing customer by offering him some incentives (Yan, Miller, Mozer, & Wolniewicz, 2001). In this competitive and volatile environment, it makes every economic sense to have a strategy to retain customers, which is only possible if the customer intention to churn is detected early enough.

There are many different reasons for customers to churn, some of them, like moving home, unstoppable, others like sudden death, undetectable. The churn prediction systems therefore should focus on detecting those churners that are deliberately moving to a competitor as these customers are most likely to leave data traces of their intent prior to churn and can be potentially persuaded to stay. This work is not concerned with the effectiveness of actual actions preventing customer churn or rescuing customers who cancelled their contract. The only concern here is the prediction of customer churn in order to provide the information about which customers are most likely to leave the service in the near future.

Churn prediction attracts recently a lot of both scientific and business attention. In the presence of large data warehouses as well as terabytes of data from Web resources, data mining techniques are increasingly being appreciated and adopted to business applications (Lemmen, 2000; Morgan, 2003), in an attempt to explain drivers of customer actions, in particular sudden falls in customer satisfaction and value ultimately leading to churn.

There is a number of churn prediction models used commercially at present, however churn is only being modelled statically by analysing event-driven customer data and running regression or predictive classification models at a particular time (Duda, Hart, & Stork, 2001) over aggregated customer data. Some improvement is obtained after segmenting customers into specific groups and dealing with different groups separately yet this segmentation only supports company’s customer relationship management (CRM) and on its own does not improve weak performance in churn prediction. In practise the most common churn management systems are even simpler as they try to devise a churn risk based on regression against available data variables. On the research arena the focus is shifted towards more complex classification and non-linear regression tech-
Related Content

Implementation of Data Mining Algorithm With R
C. Deisy and Mercelin Francis (2019). Sentiment Analysis and Knowledge Discovery in Contemporary Business (pp. 92-125).
www.igi-global.com/chapter/implementation-of-data-mining-algorithm-with-r/210966?camid=4v1a

Reducing a Class of Machine Learning Algorithms to Logical Commonsense Reasoning Operations
Xenia Naidenova (2008). Mathematical Methods for Knowledge Discovery and Data Mining (pp. 41-64).
www.igi-global.com/chapter/reducing-class-machine-learning-algorithms/26132?camid=4v1a

Federated Data Warehouses
Stefan Berger and Michael Schrefl (2010). Complex Data Warehousing and Knowledge Discovery for Advanced Retrieval Development: Innovative Methods and Applications (pp. 82-107).
www.igi-global.com/chapter/federated-data-warehouses/39589?camid=4v1a

ONTOLOGY LEARNING and the HUMANITIES
Toby Burrows (2011). Ontology Learning and Knowledge Discovery Using the Web: Challenges and Recent Advances (pp. 186-199).
www.igi-global.com/chapter/ontology-learning-humanities/53887?camid=4v1a