Case Studies in Applying Data Mining for Churn Analysis

Susan Lomax, University of Salford, Salford, UK

Sunil Vadera, University of Salford, Salford, UK

ABSTRACT

The advent of price and product comparison sites now makes it even more important to retain customers and identify those that might be at risk of leaving. The use of data mining methods has been widely advocated for predicting customer churn. This paper presents two case studies that utilize decision tree learning methods to develop models for predicting churn for a software company. The first case study aims to predict churn for organizations which currently have an ongoing project, to determine if organizations are likely to continue with other projects. While the second case study presents a more traditional example, where the aim is to predict organizations likely to cease being a subscriber to a service. The case studies include presentation of the accuracy of the models using a standard methodology as well as comparing the results with what happened in practice. Both case studies show the significant savings that can be made, plus potential increase in revenue by using decision tree learning for churn analysis.

KEYWORDS

Business Intelligence, Churn Prediction, Data Mining, Decision Trees, Knowledge Discovery

INTRODUCTION

Organizations in many different domains such as wireless telecommunication and the telecommunication industry (Huang et al., 2012, Keramati, 2014, Mahajan et al., 2015), mobile phone (Kirui et al., 2013), internet service providers (Khan et al., 2010), energy providers and other industries such as insurance, retail banking (Mutanen et al., 2006), financial services and supermarkets are having increasing difficulty in attracting and retaining customers as reported by Shandiz (2015). This is in part owing to customers being able to access information regarding brands, products and price comparisons on many internet comparison websites (Mahajan et al., 2015). The cost of acquiring new customers is higher than retaining old ones (Kirui et al., 2013) and small changes in the retention rate have been shown to have significant impact on businesses (Van den Poel & Larivie`re 2004; Larivie`re & Van den Poel 2005). For example, in the banking industry, Reichheld & Sasser (1990) and Nie et al. (2011) conclude that a bank is able to increase profits by 85% as a result of a 5% improvement on its retention rate.

Thus, a key part of any business these days is to manage customer churn; that is, to avoid losing its customers and build its customer base (Mattison 2005; Tsai & Lu 2010; Nie et al., 2011; Kirui et al., 2013).

A number of authors have advocated the use of data mining techniques to develop models to predict possible churn (Kirui et al., 2013). Several methods have been proposed including decision trees, neural nets, K Nearest Neighbour, logistic regression, random forests, SVM, linear and quadratic discriminate analysis, GA, Markov model, cluster analysis and optimization (Hadden et al., 2005; Better et al., 2008; Nie et al., 2011; Shandiz 2015). Most of these studies provide useful results on benchmark data, showing the potential for applying these methods for prediction of churn in practice. The experiences and lessons learned from applying data mining for predicting customer churn are seldom reported. This is perhaps owing to commercial confidentiality. It can also be difficult to find public datasets which can be used for churn prediction owing to business privacy and confidentiality therefore case studies presented using a variety of features are particularly useful (Kirui et al., 2013).

Thus, this paper presents two case studies in applying data mining for predicting customer churn. Although limited to two case studies, the use of non-traditional data in one of the case studies attempts to address this issue.

BACKGROUND

Churn, also known as turnover, defection or attrition is the loss of clients or customers. Many domains such as banks, mobile phone companies, internet service providers and supermarkets use churn analysis and churn rates as a key business metric as it has been shown that the cost of retaining an existing customer is less than the cost of acquiring new customers (Wei & Chiu 2002; Hung et al., 2006; Huang et al., 2012). These existing customers tend to purchase more than new customers and it is more efficient to deal with existing customers than dealing with new customers (Fornell & Wernerfelt 1987; 1988; Reichheld & Sasser 1990; Bolton 1998).

Related Content

Fuzzy-Based EOQ Model With Credit Financing and Backorders Under Human Learning
Mahesh Kumar Jayaswal, Mandeep Mittal, Isha Sangal and Jayanti Tripathi (2021). *International Journal of Fuzzy System Applications (pp. 14-36).*
www.igi-global.com/article/fuzzy-based-eoq-model-with-credit-financing-and-backorders-under-human-learning/288393?camid=4v1a

Toward Formal Verification of SDN Access-Control Misconfigurations
Amina Saadaoui (2019). *Artificial Intelligence and Security Challenges in Emerging Networks (pp. 146-161).*
www.igi-global.com/chapter/toward-formal-verification-of-sdn-access-control-misconfigurations/220550?camid=4v1a

Automated Cryptanalysis
Otokar Grošek and Pavol Zajac (2009). *Encyclopedia of Artificial Intelligence (pp. 179-185).*
www.igi-global.com/chapter/automated-cryptanalysis/10245?camid=4v1a

CoAP-Based Lightweight Interoperability Semantic Sensor and Actuator Ontology for IoT Ecosystem