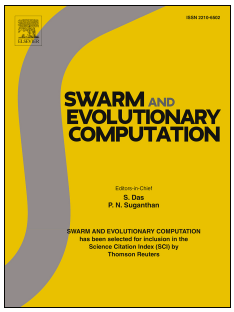
Accepted Manuscript



Profit maximizing logistic model for customer churn prediction using genetic algorithms

Eugen Stripling, Seppe vanden Broucke, Katrien Antonio, Bart Baesens, Monique Snoeck

PII: S2210-6502(16)30175-4

DOI: 10.1016/j.swevo.2017.10.010

Reference: SWEVO 336

To appear in: *Swarm and Evolutionary Computation BASE DATA*

Received Date: 4 August 2016

Revised Date: 7 September 2017

Accepted Date: 19 October 2017

Please cite this article as: E. Stripling, S. vanden Broucke, K. Antonio, B. Baesens, M. Snoeck, Profit maximizing logistic model for customer churn prediction using genetic algorithms, *Swarm and Evolutionary Computation BASE DATA* (2018), doi: 10.1016/j.swevo.2017.10.010.

Profit Maximizing Logistic Model for Customer Churn Prediction Using Genetic Algorithms

Eugen Stripling[a,*], Seppe vanden Broucke[a], Katrien Antonio[b,c], Bart Baesens[a,d], Monique Snoeck[a]

[a] *Department of Decision Sciences and Information Management, KU Leuven, Leuven, Belgium*
[b] *Department of Accountancy, Finance and Insurance, KU Leuven, Leuven, Belgium*
[c] *Section of Quantitative Economics, University of Amsterdam, Amsterdam, The Nederlands*
[d] *School of Management, University of Southampton, Southampton, UK*

Abstract

To detect churners in a vast customer base, as is the case with telephone service providers, companies heavily rely on predictive churn models to remain competitive in a saturated market. In previous work, the expected maximum profit measure for customer churn (EMPC) has been proposed in order to determine the most profitable churn model. However, profit concerns are not directly integrated into the model construction. Therefore, we present a classifier, named ProfLogit, that maximizes the EMPC in the training step using a genetic algorithm, where ProfLogit's interior model structure resembles a lasso-regularized logistic model. Additionally, we introduce threshold-independent recall and precision measures based on the expected profit maximizing fraction, which is derived from the EMPC framework. Our proposed technique aims to construct profitable churn models for retention campaigns to satisfy the business requirement of profit maximization. In a benchmark study with nine real-life data sets, ProfLogit exhibits the overall highest, out-of-sample EMPC performance as well as the overall best, profit-based precision and recall values. As a result of the lasso resemblance, ProfLogit also performs a profit-based feature selection in which features are selected that would otherwise be excluded with an accuracy-based measure, which is another noteworthy finding.

Keywords: Data mining, customer churn prediction, lasso-regularized logistic regression model, profit-based model evaluation, real-coded genetic algorithm

1. Introduction

In saturated markets such as the telephone service industry, companies constantly endeavor to identify customers who intend to voluntarily switch to a competitor. Attracting new customers in such markets is eminently challenging, and costs five to six times more than to prevent existing customers from churning [1]. However, detecting would-be churners out of typically millions of customers is a difficult task. For that reason, companies unavoidably have to rely on predictive churn models if they wish to remain competitive. As a consequence, predictive classification techniques for customer churn are increasingly researched [2]. Yet, these models often do not directly focus on the most important business requirement: *profit maximization*. Therefore, correctly identifying potential churners is one challenge; another is to also detect those who are the most profitable to the business. The ideal churn model is thus capable of effectively identifying churners and simultaneously taking profit concerns of the business into account.

*Corresponding author: eugen.stripling@kuleuven.be