Estimating Unbiased Effects of Campaigns on Long Term Churn Probability Using Survival Analysis: An Intelligent Data Mining Solution Applied to an Italian Mobile Phone Company

Furio Camillo and Ida D’Attoma

In the telecommunications service industry, the cost of obtaining new customers is more expensive than retaining existing ones. At this end, many companies used to run Churn Analysis to predict the behavior of customers who are most likely to change provided services, and to create special marketing tools or redemption strategies for them. Here, considering retrospectively such strategies as given and implemented on the basis of complaints analysis and/or customer satisfaction analysis, the aim is to understand how long is the duration of the customer relationship and how it changes after specific marketing and retention strategies. To evaluate the effectiveness of marketing or retention strategies (the treatment) on the length of the customer relationship, two non parametric Kaplan-Meier survival curves will be estimated: one for customers who received a marketing stimulus (treated) and one for customers who did not (untreated). Measuring such effects may be hindered by the consideration that selection bias might exist due to the lack of random assignment of customers to marketing stimuli. Therefore, in order to control for selection bias, the treatment effect will be measured only in groups of customers where relevant covariates (churn determinants and variables related to marketing stimulus assignment, e.g. behavioral, sociodemographics, structural covariates) are balanced among treated and untreated. At this end a three-step data mining process will be adopted.

The first step involves measuring imbalance via the Global Imbalance measure and then testing the extent to which there is imbalance in data via the multivariate imbalance test (D’Attoma and Camillo 2011). When data result unbalanced a Cluster Analysis is executed in order to enhance the plausibility of obtaining balanced groups, thereby minimizing selection bias. Next, the balance within step 2’s resulting clusters is assessed and local treatment effects within balanced groups are computed. Results related to churn behavior of customers of an Italian mobile company in 2010-2012 period will be illustrated. In particular, customers who do not churn will be considered as right-censored data. Once estimated the two survival curves in each balanced cluster, the effect will be computing by comparing the survival curve for treated and the survival curve for untreated (Figure below) using the Log-rank test that weights differences that occur earlier and later in the curve equally, and/or the Wilcoxon test that weights earlier differences higher than later differences. If according to these tests the two curves differ, an effect of the marketing stimulus on customers churn behavior exists.
Main References


Note: We ran two tests equality of the survivor functions between groups: the Log-rank (chi-square=122.08; $p < 0.01$) and the Wilcoxon (chi-square 109.67; $p < 0.01$). These both support the conclusion that the curves are different from one another.