2011

Network neighbor effects on customer churn in cell phone networks

Pavel N. Krivitsky
University of Wollongong, pavel@uow.edu.au

Pedro Ferreira
Instituto Superior TEcnico

Rahul Telang
Carnegie Mellon University

Publication Details

Krivitsky, P. N.., Ferreira, P. & Telang, R. (2011). Network neighbor effects on customer churn in cell phone networks. Symposium on Statistical Challenges in Electronic Commerce Research (pp. 1-2).

Network neighbor effects on customer churn in cell phone networks

Keywords
phone, network, cell, churn, customer, effects, networks, neighbor

Disciplines
Engineering | Science and Technology Studies

Publication Details
Krivitsky, P. N.., Ferreira, P. & Telang, R. (2011). Network neighbor effects on customer churn in cell phone networks. Symposium on Statistical Challenges in Electronic Commerce Research (pp. 1-2).

Network neighbor effects on customer churn in cell phone networks

Pavel N. Krivitsky[*†] Pedro M. A. Ferreira[*†] Rahul Telang[*]

1 Introduction

Churn in cell phone networks is as high as 25% per year. Consequently, carriers are interested in learning more about the determinants of churn. A number of interesting business opportunities arise when a quarter of the customer base changes so rapidly. However, such a high turnover of the customer base also raises concerns and the business can be seen as highly risky. Churn can be predicted with some degree of confidence by simply looking at calling patterns. However, it is important to understand the effects of social networks on churn, namely how the local network neighborhood influences the choice of network carrier. For that purpose, this paper studies how likely individuals are to churn as a function not only of their calling patterns but also of their social circle, within and outside their carrier.

2 Data

We use voice and SMS records, information on tariff plan and handset information for all customers of a major wireless carrier in an European country between August 2008 and June 2009. We aggregate the data into monthly call graphs. Our analysis focuses on prepaid mobile customers with consumer grade plans. Unlike postpaid customers, for whom churning is an explicit act of contacting the carrier and canceling service, churning for prepaid customers is often passive in nature: the customer ceases to make calls, send SMS messages, or add funds to the account. We thus follow the definition of churn suggested by the carrier and consider that a subscriber churns after having made no calls or sent no SMS messages for three consecutive calendar months. Empirically, we find that the number of subscribers that cease to use the service for three months and then resume is negligible.

[*]iLab, H. John Heinz III College, Carnegie Mellon University

[†]Institute for Systems and Robotics, Instituto Superior Técnico

3 Methods

For each subscriber, a full history of service plans is available, and, thus, the subscriber's initial time of subscription is known. However, according to the churn criteria used, for those subscribers who were still active after the end of March 2009, the ultimate churning time is unknown. We thus take a survival analysis approach, defining the survival time of a subscriber to be the number of calendar months between the initial subscription and the last month of activity and using the Cox proportional hazard modeling framework, with frailty, to account for individual heterogeneity, [1, pp. 296–308] and time-varying covariates. More concretely, we model the hazard of the current month being the last month of subscriber's activity given the subscriber's activity during the prior month: the hazard of subscriber i churning t months after subscribing is modeled as

$$h(t, x_i, x_i^{(t-1)}, \beta, \beta^{(-1)}) = z_i h(t) e^{x_i \cdot \beta + x_i^{(t-1)} \cdot \beta^{(-1)}},$$

where z_i are Gamma-distributed individual subscriber effects, x_i are the non-time-varying covariates for subscriber i, $x_i^{(t)}$ are i's time-varying covariates t months after i's initial subscription, and β and $\beta^{(-1)}$ are their respective coefficient.

Although the starting date for each subscriber is observed, different subscribers have different tenures at the time the period of observation of their voice and SMS activity begins. This means that the different subscribers' tenures are "shifted" relative to each other. This introduces two additional challenges. First, those subscribers who churned before the period of observation cannot be modeled, and are thus left-truncated [1, p. 228]. This does not preclude the analysis, but may make it much more sensitive to the proportionality of hazards assumption. We address this by fitting the models to two versions of the datasets: one, a sample from those subscribers of interest who were active at any time from September 2008 through March 2009 ("full"); the other, a "short" sample from only on those subscribers who became customers after August 2008.

Second, because there is likely to be a seasonal effect on churn (e.g., some firms run promotions during certain months — most notably during Christmas time), the analysis must control not only for the time since subscription but also for the calendar time. We address this by adding calendar month as a categorical predictor.

Of primary interest in our analysis is the effect of the subscriber's immediate social neighborhood. In particular, does having subscribers of other carriers as frequent contacts increase an individual's propensity to churn? We call a telephone number j a *mutual SMS neighbor* of a subscriber i in a given month if i had sent at least one message to j and j had sent at least one message to i in that month. For each subscriber, we include the number of mutual SMS neighbors within and outside the carrier's network as a covariate. Because, unlike SMS messages, telephone communication is bidirectional during a call, we use three definitions for voice neighbors: three thresholds — 1 call, 3 calls, and 5 calls — of calls between i and j (initiated by either individual) for j to be considered an *n-call neighbor* of i. For each of these thresholds, we fit a model with the number of n-call neighbors as a covariate.

To control for other factors likely to affect a subscriber's propensity to churn — such the overall level of usage — we also include tariff plan information and linear and quadratic effects of number of calls made, number of calls answered, airtime of calls made, airtime of calls answered, SMS messages sent, SMS messages received, and number of line services.

4 Results

We fit the three survival models to two random samples of 100,000 subscribers, drawn according to the schemes described above. After controlling for the other variables, number of mutual SMS neighbors inside and outside the network in the prior two months has a highly significant and consistent effect on propensity to churn. Specifically, in all the models, the coefficients on the number of mutual SMS out-of-network neighbors were positive and significant (i.e., increasing the churn hazard), while the corresponding coefficients on the in-network neighbors were all negative and significant. Contrasting the coefficients, we find that for the "full" sample, switching a mutual SMS neighbor from in-network to out-of-network and keeping all other predictors fixed increases the predicted hazard of the subscriber churning in the subsequent month by 34%. For the "short" sample, the hazard is increased by 24%. (P-val. < 0.01 for all coefficients and contrasts given here.)

The results for voice neighbors are less consistent. Indeed, we find that having more 1-call, 3-call, or 5-call neighbors either in-network or out-of-network either reduces the hazard of churn or has no statistically significant effect. However, there is a pattern to the contrasts between these coefficients. In the "full" sample, the effect of switching a 1-call neighbor from in-network to out-of-network reduces the probability of churn by 4.2% (P-val. < 0.01), for 3-call neighbors, the effect is not statistically significant (P-val. $= 0.7$), while for 5-call neighbors, such a switch increases the probability of churn by 2.4% (P-val. $= 0.04$). When only the subscribers who had subscribed during the period of observation are considered, the same general trend appears to hold, but the 5-call neighbor contrast is not statistically significant at $\alpha = 0.05$.

5 Discussion

These results suggest an unambiguous effect of effect of SMS message network neighborhood on churn over and above usage level alone. The pattern in voice effects suggests that stronger network neighbors might have greater effect on churn, although it is not clear why out-of-network weak neighbors consistently appear to reduce the hazard of churn by more than in-network neighbors. (This result holds even if mutual SMS neighbors are excluded from the model.) We have replicated the above analysis, also breaking the control variables down by in-network and out-of-network, with similar results. This is not sufficient to draw a causal conclusion [2, and others]: at least some of this effect may be due to time-varying confounders such as advertising campaigns, but the results are encouraging to delve further into how customers make decisions about carriers from whom they buy service.

References

[1] D. W. Hosmer, S. Lemeshow, and S. May. *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 2nd edition, 2008.

[2] C. R. Shalizi and A. C. Thomas. Homophily and contagion are generically confounded in observational social network studies. Technical report, Department of Statistics, Carnegie Mellon University, Apr. 2010.