User Modeling for Telecommunication Applications: Experiences and Practical Implications

Heath Hohwald, Enrique Frias-Martinez, and Nuria Oliver

Data Mining and User Modeling Group
Telefonica Research, Madrid, Spain
{heath,efm,nuriao}@tid.es

Abstract. Telecommunication applications based on user modeling focus on extracting customer behavior and preferences from the information implicitly included in Call Detail Record (CDR) datasets. Even though there are many different application areas (fraud detection, viral and targeted marketing, churn prediction, etc.) they all share a common data source (CDRs) and a common set of features for modeling the user. In this paper we present our experience with different applications areas in generating user models from massive real datasets of both mobile phone and landline subscriber activity. We present the analysis of a dataset containing the traces of $50,000$ mobile phone users and $50,000$ landline users from the same geographical area for a period of six months and compare the different behaviors when using landlines and mobile phones and the implications that such differences have for each application. Our results indicate that user models for a variety of applications can be generated efficiently and in a homogeneous way using an architecture based on distributed computing and that there are numerous differences between mobile phone and landline users that have relevant practical implications.

1 Introduction

User Modeling is a key process in a wide variety of (telco) telecommunication applications in which knowledge of individual users is key for providing a better service and anticipating user needs. The most relevant applications include: (1) churn prediction, i.e. the ability to anticipate users that are at risk of leaving the company, (2) information spreading processes, such as viral and targeted marketing, which include a variety of techniques to spread information in the network and the ability to identify key users that can influence others in their decision making process, (3) fraud detection, which focuses on identifying users that will exhibit fraudulent behavior, and (4) network design and planning, which seeks to adapt and plan a network to meet the needs of the users and the design of pricing plans.

Although these applications are very different in nature they typically generate user models from a common data source. The features of the different user models frequently overlap and the architecture used to generate the models can be shared. Regarding data sources, CDRs (Call Detail Records) are used as a primary source of information for constructing user models for telco applications since they implicitly contain the behavior of each customer, from calling patterns, to consumption, terminal changes or characteristics of the social network. In some cases other extra information of each customer, such as gender, can be used. As for the dimensions used for the user models, in general there is a set of features, such as total talk time or total degree, that are relevant for a wide variety of applications. These common factors imply that the same architecture can be used for generating user models for a variety of applications.

Telco applications can be divided into two main areas: mobile and landlines. While mobile phones are in widespread usage and are typically used by just one individual, the number of landlines is much smaller and their use is typically shared by more than one individual. The user models generated for both cases use the same set of features, although the relative importance and implications of each feature differs.

In general the architectures used for generating user models for telco applications have to be very data intensive in order to process the amount of data available (typically several months) for all the customers (typically several million). The main differences between applications are not so much in the way user models are generated or in the features of the user models, but in the training sets used to construct the classifiers, i.e. while the training set for churn prediction will include users that have churned, for fraud prediction they will include users that have committed fraud. This implies that the same architecture can be used to generate different user models for different applications, and that there is no need for ad-hoc solutions.

In this paper we present our experience in generating user models from real CDR traces for telco applications. Also we compare the differences between landline and mobile phones for each feature and the practical implications that those differences have. The rest of the paper is organized as follows: after presenting the related work, we detail the construction of user models and typical features used for telco applications. Section 4 presents the Methodology for User Modeling and Section 5 the lessons learned and the implications for different applications of the features studied. We conclude in Section 6.

2 Related Work

The literature reports a wide variety of studies related to telco applications. Most of the work has focused on studies using mobile phone data [18, 6, 13, 15], while landline data has received less attention [5, 1]. Churn prediction algorithms have been implemented for landlines[16] and mobile phones [9, 6, 2]. Traditionally, churn prediction has been solved with classification techniques that predicted in which group (churner or non-churner) a given user was included. User models were constructed using implicit information provided by CDR data such as calling patterns [19] or social network patterns [6]. The techniques used for creating the classifier encompass typical machine learning techniques such as: neural networks [3], classification trees [4], SVM [2] and genetic algorithms [9]. Information spreading algorithms originally appeared in social sciences [10] and are based on the idea of using a social interaction network to model the flow of information and influence. The concept groups a variety of algorithms that model the pervasive word of mouth behavior and are typically based on the spreading activation method used in cognitive psychology. These family of algorithms have been successfully used in a variety of telco applications, including viral marketing [17], churn prediction [6], and modeling of trust [22]. Fraud detection in the telco context aims at detecting individuals that acquire a mobile phone and do not intend to pay their contract [8]. Typical approaches focus on classifying users according to their level of risk by calculating deviations from standard behaviors [20, 11]. Telco user models have also been effectively used for the improvement of the network infrastructure, including the design of pricing plans, an application where mobility data has proven extremely relevant. For example [21] modeled number of calls, number of cells visited and the entropy of user locations for voice, data and SMS in order to improve paging efficiency in cellular networks.

Our work, when compared to previous approaches, presents three main novel elements: (1) the techniques used for each solution are typically developed adhoc, but we consider that although the applications are very different, the fact that they share the data source and a lot of dimensions implies that the same architecture can be used to generate the user models needed, (2) in general previous approaches use a limited number of users, while we consider one of the key challenges of user modeling is going to be the ability to obtain conclusions from massive datasets, and (3) we present the first analysis of the differences between landlines and mobile phones and the implications that those differences have for telco applications.

3 Generating User Models for Telco Applications

3.1 Telco Data Acquisition

Mobile phone networks are constructed using base transceiver stations (BTS) that are in charge of communicating mobile phones with the network. The area covered by a BTS is called a cell. Call Detail Records (CDRs) are generated when a mobile phone connected to the network makes or receives a phone call or uses a service (SMS, MMS, etc.). In the process, the information regarding the connection is stored in the form of a Call Detail Record, which includes the originating phone number, the destination phone number, the time and date of the call, the total length of the call and the BTS used for the communication. The originating and destination numbers are encrypted to preserve privacy. The BTS gives an indication of the geographical position of the user, but no indication of the position of a user within the cell is known. CDR data for landline subscribers is acquired in a similar fashion but without the need for a BTS. Typically CDRs for a given period of time are stored in more than one file, for example one file per day, which facilitates the generation of user models when using data-driven architectures (See section 3.3).

3.2 Features of Telco User Models

In this section we present a set of features that have been found to be generally useful for generating both landline and mobile user models across a range of applications.

Total Number and Total Duration of Calls Two of the most basic metrics that can be computed for each user are the total number of calls and the total talk time over a specified time period. The number of calls and total talk time can each be further restricted according to direction of call, where each subscriber's incoming and outgoing calls are considered separately. From an application perspective, variations of these features are very relevant, for example the ratio between national and international calls or between calls made within the provider and outside the provider are very relevant for churn [9, 2] and fraud detection[20, 11]. Also these two variables are relevant to viral marketing as users that have a lot of connections are more capable of spreading information [10]. As for network design, these are key features used to balance the network [21].

Calling Behavior for Each Day While features such as total number of calls and total duration capture a user's aggregate activity level, a vector of temporal features can be used to capture the variation in calling behavior during the course of the day or week. Considering first daily behavior, for each user two vectors of length seven record the total number of calls and total talk time for each day of the week. The same features can be computed for the reciprocal call CDR data sets. A day-by-day comparison between landline and mobile reciprocal call data is indicative of what day of the week each set of users tends to speak with members of their social circle and can be an important factor in targeted advertising campaigns. Also this information is very relevant for fraud as it is used to generate the user model that describes normal behavior [20] and churn[19].

Calling Behavior for Each Hour of the Day Similar to the features that segment activity by day of the week, it is possible to calculate the number of calls and total talk time for each user based on the time that each call was initiated. Typically, the time intervals considered are 24 one-hour long bins beginning at the start of each hour. For each user, two vectors of length 24 can be constructed in order to capture the total number of calls and talk time for each hour of the day, aggregating over all days in the data. The vectors provide insight into understanding what time of the day each user tends to have most of their calls and speak the most. The percentage of calls and talk time coming from *reciprocal* talk partners indicates the time of day when each user is most likely to be speaking with members of their social circle, a key element for designing viral marketing campaigns. Considering the aggregate results for the entire population is useful for network planning, since the network operator must plan for the different peaks in usage for landline and mobile networks [21]. As in the previous case, this information is very relevant for fraud detection[20] and churn[19].

Social Network Features The concept of degree is one of the fundamental metrics in social network analysis. A graph $G_D = (E, V)$ that represents the social network of the callers present in the data may be derived from CDR data D. Each node $v \in V$ corresponds to a different phone number and each directed edge $e = (v_1, v_2) \in E$ corresponds to a call from node v_1 to node v_2. In this context, the degree of a node v, denoted $Deg(v)$, corresponds to how many distinct talk partners subscriber v has and is given by the number of edges incident with node v. The in-degree of a node v corresponds to the number of distinct individuals that call v while the out-degree is given by the number of distinct individuals called by v. Reciprocal degree of a node v corresponds to the total number of edges incident with ode v in the reciprocal graph and is a measure of the total number of talk partners in a user's true social circle. The higher the degree, the larger the social circle. It is important to maintain customers with large social circles since they can exert influence on a large number of other subscribers, potentially causing them to churn [14]. Recently this information has also been included in churn prediction models [6]. The reciprocal degree is key element for viral marketing, an in general for diffusion information processed, because provides a way of identifying strong ties [10].

3.3 Construction of User Models

As illustrated above, there are a large number of features that can be calculated from a given set of CDR data for either landline or mobile subscribers and that are relevant for a variety of applications. The construction of telco user models is complicated by the fact that often CDR records usually contain several months of data with hundreds of millions of records for tens of millions of users. Rather than constructing each user model for each application area, we have developed ARBUD [12], an terabyte architecture for automating the user model construction process that is based on a distributed computing paradigm and typically run on a computer cluster. One of the components of ARBUD is a library containing reusable modules for constructing different features of a user model in an efficient way. All of the features mentioned in section 3.2 have been added as modules in the ARBUD library, typically using the MapReduce programming paradigm [7]. A metamodel is used for specifying the desired features, location of the data and any other relevant parameters and ARBUD then interprets the metamodel and constructs the desired user models.

4 Experimental Setup

In order to compare residential mobile and landline users, two random samples of $50,000$ mobile and $50,000$ landline subscribers were drawn from the same metropolitan area. The sample of landline users is denoted as S_L and mobile users as S_M. Any overlap between S_L and S_M was arbitrary and not identifiable. A set of CDR data was obtained for the subscribers in S_L and S_M during the same six-month period. The CDR data associated with S_L and S_M are denoted by D_L and D_M and their size by $|D_L|$ and $|D_M|$, respectively. All calls for S_M were recorded, even when the subscriber left the geographic region from which the sample was drawn. The total number of calls (in millions) was $|D_L| = 50.3$ and $|D_M| = 41.8$.

The information present in each CDR includes the encrypted originating phone number, the encrypted destination phone number, the duration of the call in seconds, and the time and date when the call originated. The sets of all reciprocal calls made and received by the subscribers in S_L and S_M are denoted by $D_{L,R}$ and $D_{M,R}$ (and their sizes by $|D_{L,R}|$ and $|D_{M,R}|$) respectively. The total number of reciprocal calls (in millions) was $|D_{L,R}| = 27.3$ and $|D_{M,R}| = 29.9$.

ARBUD was used to build user models from both the landline and mobile datasets for the users in S_L and S_M, including all the features mentioned in Section 3.2. The construction process was carried out on a cluster with 5 machines, each with 16 GB of RAM, 4 hard drives each with 1 Terabyte storage capacity, and 4 quad core processors. The nodes were all connected with a fast gigabit network switch. Both models were constructed in less than 24 minutes.

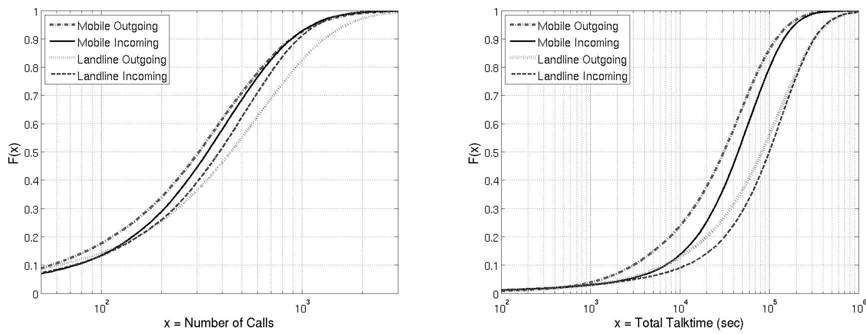
5 Results and Discussion

Using the user models generated in Section 4, this section studies the features presented, contrasting the differences between landline and mobile users and their implications from a practical perspective.

5.1 Total Number and Total Duration of Calls

Figure 1a depicts the empirical CDFs (Cumulative Distribution Function) for four different distributions: the number of incoming and outgoing calls for landline and mobile subscribers. Looking at the median for each distribution, which corresponds to the horizontal line where $F(X) = 0.5$, we observe that of the four distributions, the lowest number of calls corresponds to outgoing calls from mobile subscribers with about 300 outgoing calls over the 6-month period, or a little less than 2 outgoing calls a day. For the region in question, mobile subscribers only pay for calls they place, so unsurprisingly mobile users make fewer calls than they receive. Landline users make and receive more calls than mobile users. In the case of the 10% of users from each population that have the most calls, three of the distributions have a similar number of calls, while the distribution for landline outgoing calls has a higher value, with the top 10% of the landline sample making at least 1500 calls during the 6-month period. This information is very relevant to offer and design new plans to subscribers. Also the top users of the distribution are key individuals as they can play a relevant role in the information spreading process needed for viral marketing and churn (other features such as degree are also related).

Figure 1b represents the empirical CDFs for the total talk time. Four different distributions are depicted: the total talk time of incoming and outgoing calls for mobile and landline users. In this case, for the median subscriber the smallest amount of talk time corresponds to mobile subscribers making calls (outgoing), followed by mobile subscribers receiving calls (incoming) and landline subscribers making calls (outgoing). Landline subscribers receiving calls (incoming) have the highest talk time of all groups, even though the largest number of calls corresponded to outgoing calls from landline subscribers. Looking again at the



(a) Number Incoming and Outgoing Calls

(b) Total Talktime

Fig. 1: CDFs for Number of Incoming and Outgoing Calls and Total Talktime

median of each distribution, a total talk time of about $30,000$ seconds (8.3 hours) is seen for outgoing mobile calls, while a talk time of about $100,000$ seconds (27.7 hours) is seen for incoming landline calls: at the median of each distribution, landline subscribers spend more than 3 times more talking on the phone when receiving calls than mobile subscribers when making calls.

Figure 2 presents the CDFs of average call duration of S_L and S_M users for incoming and outgoing calls. The median value for the average call duration for incoming landline calls is more that twice the call duration of outgoing mobile calls, likely indicating sensitivity to different pricing structures. Mobile users in the top 10% of talk time are seen to talk for more than 30 times as much as the bottom 10%, so there is very heterogeneous behavior which helps explain the success of numerous pricing plans.

5.2 Calling Behavior for Each Day of the Week

The total number of calls (the sum of all the made/received calls) by all subscribers in the landline (S_L) and mobile (S_M) samples aggregated over all user models built for D_L and D_M for each day of the week are shown in Figure 3a.

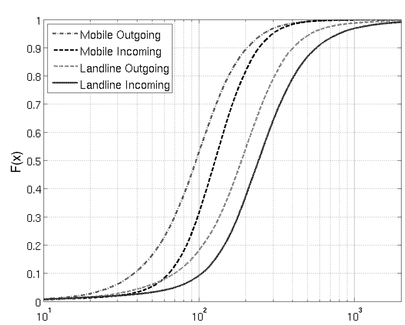
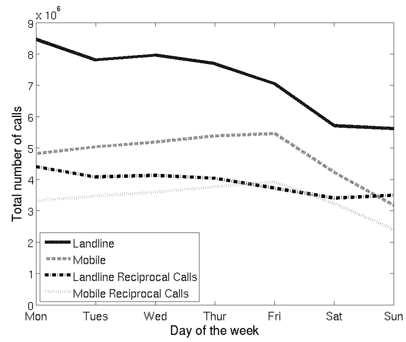
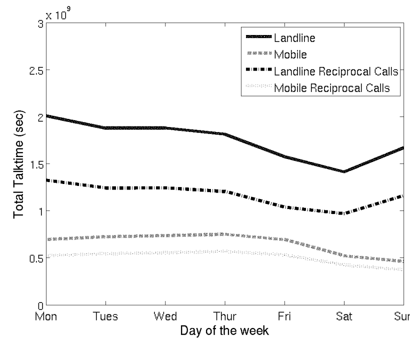


Fig. 2: CDFs for Average Call Duration for Incoming and Outgoing Calls



(a) Number of Calls



(b) Total Talktime

Fig. 3: Number of Calls and Total Talktime by Day of the Week

Since both the mobile and landline samples are for $50,000$ subscribers during the same time period, the aggregate number of calls and talk time for each day of the week can be directly compared on a day-by-day basis. Figure 3a also depicts the total number of calls with only reciprocal talk partners for each day of the week, obtained by summing for each day over $D_{L,R}$ and $D_{M,R}$. It can be seen that landline users make and receive more calls than mobile users for every day of the week. Interestingly, Monday is the day with most calls for landline users, while Friday is the day with most calls for mobile users. Both populations make fewer calls on the weekend. When considering only reciprocal calls, the differences between landline and mobile users decrease significantly and there are more mobile than landline phone calls on Fridays. These results are indicative of the culture of the sampled region, where Friday is the day when mobile users are most likely to make plans with their social circle and Monday is the day when landline subscribers make most non-social calls. These results can help inform a targeted advertising campaign, suggesting, for example, that socially oriented advertising directed to mobile subscribers may be best received on Fridays.

When looking at total talk time for each day of the week (Figure 3b), the results are notably different. While Sunday is the day with the fewest calls for both landline and mobile users (see Figure 3a), Saturday is the day with the least talk time in both cases. On Sunday, both populations have relatively few calls but calls tend to last longer, as evidenced by a rise in talk time despite a drop in the number of calls. Mobile users spend less time on the phone than landline users, considering both the full set of CDRs (D_M and D_L) and the reciprocal CDRs ($D_{M,R}$ and $D_{L,R}$). These differences indicate that for fraud detection the definition of a standard behavior is dependent on the type of communication (landline or cell phone) and of cultural elements.

Looking at what percentage of all calls for each day of the week are reciprocal calls (Figure 4), it can be seen that, both landline and mobile users, the weekend tends to be the time when users are most likely to have calls with their reciprocal partners. These results indicate that for the geographical region under consideration, the weekend is the time when the highest percentage of phone usage is dedicated to speaking with one's social circle.

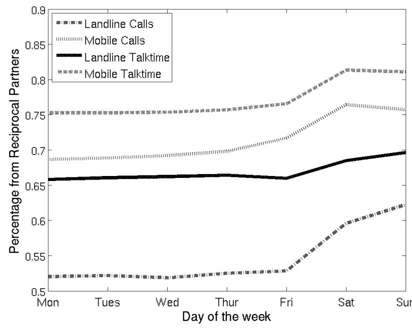


Fig. 4: Percentage of Calls and Talktime from Reciprocal Partners for Different Days of the Week

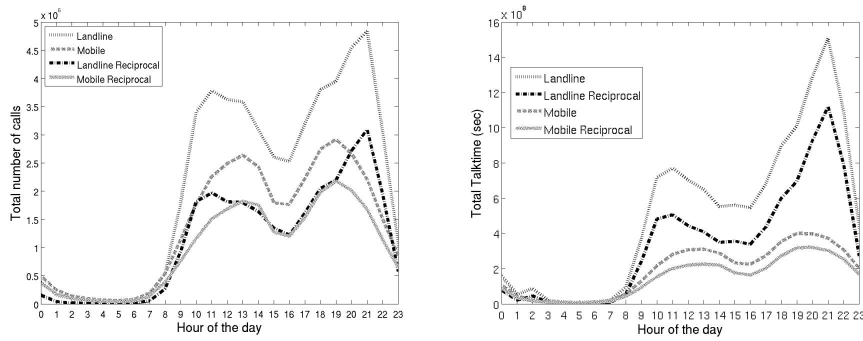
5.3 Calling Behavior for Each Hour of the Day

The number of calls as well as the number of reciprocal calls for landline and mobile subscribers for each hour of the day are depicted in Figure 5a. Both landline and mobile users show a similar trend, with very few calls in the early morning (from 0.00 to 9.00)and a significantly larger number of calls during the day (from 9.00 to 22.00). While the global maximum for mobile subscribers is at 19.00 to 20.00, the maximum for landline users takes place 2 hours later (from 21.00 to 22.00). The daily rhythm observed is indicative of the culture of the region sampled and would likely be different for other cultures. This information is also key for defining standard behavior in fraud detection and complements the information of the previous section.The fact that both groups have similar rhythms suggests that mobile and landline phones are not competing technologies, but rather complement each other. It can also be observed that landline users place and receive more calls than mobile subscribers at every hour of the day. This may be a result of the generally higher tariffs for mobile phones or the fact that landlines tend to be shared among several users while mobile phones tend to be exclusively used by a single person. Similar behavior is found when considering calls to reciprocal call partners.

The total talk time for each hour of the day (Figure 5b) follows a similar two-peak pattern. However, the peaks in mobile talk time are significantly less pronounced than the peaks in landline talk time and the amount of (reciprocal and non-reciprocal) landline talk time is significantly larger than the mobile talk time: mobile users seem to use their devices uniformly during the day, particularly in terms of talk time, whereas landline users tend to talk longer in the evening. Figures 5a and 5b indicate also that planning for peak usage in mobile and landline networks requires focusing on different time windows.

5.4 Size of Each User's Social Circle

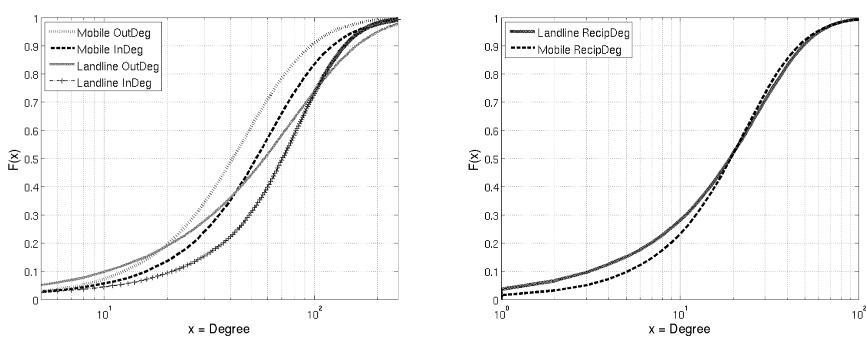
The analysis of number of calls and talk time reveals patterns of behavior at the individual level. However, they do not capture much about the social networks



(a) Number of Calls (b) Total Talktime

Fig. 5: Number of Calls and Total Talktime by Time of Day

of each individual. In order to see how landline and mobile users may differ in the size of their respective social circles, in and out degrees were calculated for all landline and mobile subscribers, resulting in four empirical distributions, which are plotted in Figure 6a. Looking at the median of each distribution, the smallest degree is for outgoing calls from mobile subscribers, with the median mobile subscriber making calls to 40 different subscribers, while the median landline subscriber receives calls from 70 different subscribers.



(a) In and Out Degree (b) Reciprocal Degree

Fig. 6: CDFs for In, Out, and Reciprocal Degree

The shortcoming of focusing on in or out degree computed over the full data set is that it does not take into account the *strength* of social connections. If degrees are calculated over the reciprocal data sets, however, a more accurate picture of the true size of each user's social circle is obtained. Figure 6b depicts empirical CDFs for reciprocal degrees calculated over the reciprocal CDR data sets $D_{L,R}$ and $D_{M,R}$. Note that when calculating degree over the reciprocal CDR data sets, it no longer makes sense to speak of in or out degree but of *reciprocal* degree. In addition, 3,956 landline and 934 mobile subscribers from S_L and S_M respectively did not have any reciprocal relationships and were not included in the empirical CDFs depicted in Figure 6b. As seen in the Figure, the distributions are significantly different than those seen in Figure 6a. In the reciprocal case, the median of the landline and mobile phone degree distributions are almost identical and lower than the degrees shown in Figure 6a, even though landline users tend to make/receive more calls and a landline is typically used by multiple individuals. In our dataset, the median size of the social circle –inferred from the reciprocal degree distributions– of landline and mobile users is 20. The top users of the reciprocal degree distribution are very important because with a social circle of size 50 or more each user can exert large influence on other subscribers, influencing their propensity to churn and/or for implementing viral marketing campaigns.

6 Conclusions and Future Work

Many different application areas in the telco domain rely upon user models. In this paper, we presented a set of features common to many telco applications and indicated why certain features are particularly relevant in certain applications. Because the main telco applications use the same data source (CDR), and the same set of features to construct user models, we proposed a general data-driven architecture to build user models for any telco application. In order to illustrate typical behavior found when building user models for real datasets, we analyzed and compared the behavior of the mobile and landline traces of $50,000$ anonymized individuals in a metropolitan area during 6 months. In our analysis, three factors were taken into account: (1) aggregate individual behavior, (2) temporal behavior, and (3) social network. The analysis identified, among others, that usage patterns depend on the time of day and the day of week, with landline users making/receiving more calls and talking longer than mobile users. Also, while mobile users made fewer calls and talked less than landline users, they tend to use their phones more often to communicate with people in their social circle, where reciprocal degree proved a useful concept for this analysis.

For future work, we plan to carry out similar analysis for different geographic regions in order to determine how generalizable the results shown here are and to what extent cultural differences affect the behavioral patterns of landline and mobile phone users.

References

1. J. Abello, P. Pardalos, and M. G. C. Resende. On maximum clique problems in very large graphs. pages 119–130, 1999.
2. C. Archaux, H. Laanaya, A. Martin, and A. Khenchaf. An SVM based churn detector in prepaid mobile telephony. In *Int. Conf. Information & Communication Technologies (ICTTA)*, pages 19–23, 2004.
3. W. Au, K. Chan, and X. Yao. A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Trans. Evol. Comp.*, 7(6):532–545, 2003.
4. L. Bin, S. Peiji, and L. Juan. Customer Churn Prediction Based on the Decision Tree in Personal Handyphone System Service. In *2007 Int. Conf. Service Systems and Service Management*, pages 1–5, 2007.

5. C. Cortes, D. Pregibon, and C. Volinsky. Communities of interest. *Lecture Notes in Computer Science*, 2189:105–114, 2001.
6. K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, and A. Joshi. Social ties and their relevance to churn in mobile telecom networks. In *EDBT '08*, pages 668–677. ACM, 2008.
7. J. Dean and S. Ghemawat. Mapreduce: Simplified data processing on large clusters. pages 137–150, 2004.
8. P. Estévez, C. Held, and C. Perez. Subscription fraud prevention in telecommunications using fuzzy rules and neural networks. *Expert Systems with Applications*, 31(2):337–344, 2006.
9. J. Ferreira, M. Vellasco, M. Pacheco, and C. Barbosa. Data mining techniques on the evaluation of wireless churn. In *ESANN2004 European Symposium on Artificial Neural Networks*, pages 483–488. Citeseer, 2004.
10. J. Goldenberg and B. Libai. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 12(3):211–223, 2001.
11. C. Hilas and J. Sahalos. User profiling for fraud detection in telecommunication networks. In *5th Int. Conf. technology and automation*, pages 382–387, 2005.
12. H. Hohwald, E. Frias-Martinez, and N. Oliver. ARBUD: A Reusable Architecture for Building User Models from Massive Datasets. UMAP 2010 Submitted, 2010.
13. A. A. Nanavati, S. Gurumurthy, G. Das, D. Chakraborty, K. Dasgupta, S. Mukherjea, and A. Joshi. On the structural properties of massive telecom call graphs: findings and implications. In *CIKM '06: Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 435–444, New York, NY, USA, 2006. ACM.
14. How to generate customer loyalty in mobile markets. Technical report, Nokia Siemens Networks, March 2009.
15. J. P. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, and A. L. Barabasi. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, May 2007.
16. J. Qi, Y. Zhang, H. Shu, Y. Li, and L. Ge. Churn prediction with limited information in fixed-line telecommunication. In *Proc. 5th Int. Symp. Communication Systems Networks and Digital Signal Processing*, pages 423–426, 2006.
17. M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *Proc. 8th ACM SIGKDD*, page 70. ACM, 2002.
18. M. Seshadri, S. Machiraju, A. Sridharan, J. Bolot, C. Faloutsos, and J. Leskovec. Mobile call graphs: Beyond power-law and lognormal distributions. In *KDD '08*, pages 596–604, 2008.
19. C. Wei and I. Chiu. Turning telecommunications call details to churn prediction. In *Expert Systems with Applications 23*, pages 4103–112, 2002.
20. D. Xing and M. Girolami. Employing Latent Dirichlet Allocation for fraud detection in telecommunications. *Pattern Recognition Letters*, 28(13):1727–1734, 2007.
21. H. Zang and J. C. Bolot. Mining call and mobility data to improve paging efficiency in cellular networks. In *MobiCom '07*, pages 123–134, New York, NY, USA. ACM.
22. C. Ziegler and G. Lausen. Spreading activation models for trust propagation. In *Proc. IEEE Int. Conf. on e-Technology, e-Commerce, and e-Service*, pages 83–97. Citeseer, 2004.