Pervasive and Mobile Computing 41 (2017) 166–178

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

Data analysis and call prediction on dyadic data from an understudied population

Mehwish Nasim [a,b,*], Aimal Rextin [c], Shamaila Hayat [c,d], Numair Khan [e], Muhammad Muddassir Malik [f]

[a] Department of Computer & Information Science, University of Konstanz, Germany
[b] School of Mathematical Sciences, The University of Adelaide and ARC Centre of Excellence for Mathematical and Statistical Frontiers, Australia
[c] Department of Computer Science, COMSATS Institute Of Information Technology, Islamabad, Pakistan
[d] Department of Computer Science, The University of Poonch, Rawlakot, Azad Jammu and Kashmir
[e] Department of Computer Science, Brown University, USA
[f] National University of Sciences and Technology, Pakistan

ARTICLE INFO

Article history:
Received 28 March 2017
Received in revised form 7 July 2017
Accepted 12 August 2017
Available online 24 August 2017

Keywords:
Smartphone
Call-logs
Call prediction
Temporal regularity

ABSTRACT

In this paper we predict outgoing mobile phone calls using machine learning and time clusters based approaches. We analyze to which extent the calling activity of mobile phone users is predictable. The premise is that mobile phone users exhibit temporal regularity in their interactions with majority of their contacts. In the sociological context, most social interactions have fairly reliable temporal regularity. If we quantify the extension of this behavior to interactions on mobile phones we expect that pairwise interaction is not merely a result of randomness, rather it exhibits a temporal pattern. To this end, we not only tested our approach on an original mobile phone usage dataset from a developing country, Pakistan, but we also analyzed the famous Reality Mining Dataset and the Nokia Dataset (from a European country), where we found an equitable basis for comparison with our data. Our original data consists of 783 users and more than 12,000 active dyads. Our results show that temporal information about pairwise user interactions can predict future calls with reasonable accuracy.

1. Introduction

Social networks are made up of a set of social entities (people, actors, organizations etc.) and social relations (friendship, kinship, etc.), between those entities. Social relations consists of persistent relations such as friendship and instantaneous relations (interactions) such as talk to, joint participation in an event, extend help to, etc. Seemingly autonomous individuals and organizations in a social network are, in fact, embedded in social relations and interactions. Massive amount of relational event data is generated by social interactions. Such data, as proxy of human relationships is helpful in understanding and predicting behavior of individuals such as influence, activity bursts, buying habits etc.

Mobile phones are the most common medium for social interactions. In America alone, there are almost 1.3 billion mobile communication events daily [1]. Because of mobile phone's portable nature, a communication event can take place in a variety of situations and one can assume very little about the context of a call.

* Corresponding author at: School of Mathematical Sciences, The University of Adelaide and ARC Centre of Excellence for Mathematical and Statistical Frontiers, Australia
E-mail addresses: mehwish.nasim@adelaide.edu.au (M. Nasim), aimal.rextin@comsats.edu.pk (A. Rextin), shamailahyt@yahoo.com (S. Hayat), nkhan6@cs.brown.edu (N. Khan), muddassir.malik@seecs.edu.pk (M.M. Malik).

http://dx.doi.org/10.1016/j.pmcj.2017.08.002
1574-1192/

This work is motivated by two factors: (1) User interactions on social media such as on smartphones carry important information about the underlying social dynamics. Discovering this knowledge is challenging because of the frequency and versatility of use of smartphones. (2) Further, the same two factors also necessitate an extremely efficient call-making interface design.

Contributions:

1. Most of the call-logs data analyzed in the literature is from developed countries [2]. We collected an original mobile phone usage dataset of 783 users with 229,450 communication events from an understudied population: Pakistani mobile phone users. This is one of the potential strengths of our work.
2. We explore temporal homogeneity/non-homogeneity in mobile phone calls, in order to predict future communication events between pairs of individuals. We perform a study of possible features in time series analysis that are useful in call prediction. Using actual call logs we show that majority of users are not optimally served by existing calling applications such as call logs. Further, we also test the hypothesis that, majority of caller–callee interactions display temporal regularity through a statistical measure called autocorrelation. We then propose a machine learning call prediction method based on temporal regularities between ego-alter pairs and perform experiments on both the collected data as well as on the famous Reality Mining Dataset [3] to demonstrate applicability of our methods for predicting future calls.
3. Further, we show that most ego-alter pairs call around the same time and use this observation to propose a call prediction algorithm and compare it with algorithm proposed by Stefanis [4] on their dataset [5].

2. Related work

Temporal regularity can be observed in time variation of activity on online social networks such as YouTube, Twitter and Slashdot, and also in frequency of edits made on Wikipedia [6–8]. Activity on twitter in various languages shows that circadian patterns exist for tweets all around the world [9]. Temporal interactions have been used to study human behavior, for instance, commenting behavior of Facebook users (a consequence of social selection or social influence effects) [10]. Temporal interactions have also been used to predict links in social networks [11–13].

Call log data has been shown to hold significant potential of providing insights into the underlying relational dynamics of societies, evolution of relationships over time and, in the absence of survey data, the quantification and prediction of social network structures [14]. Data of calling patterns has been used to infer friendships relations and uncover individual and collective human dynamics [3,14–18]. Call-volume data has been used to explore whether the distribution of calls in an urban population follow routine patterns or not, and whether the variation of such patterns in different parts of the city can be explained [19]. Inspired by effective studies on calling patterns, researchers have devised several call prediction models. In [20], authors predicted the outgoing and incoming calls on Reality Mining dataset [3] based on most recent calling data. Out of the 94 datasets, they used a small subset of 30 users for performance evaluation. Barzaiq et al. [21] modeled the historic call patterns of users and achieved a 35% accuracy for call prediction on synthetic data. Haddad et al. [22] discuss a probabilistic model that uses call frequency to predict incoming and outgoing calls for each individual contact.

Recent studies of human behavior indicate that the timing of communication events is characterized by long dormant periods interspersed with bursts of high activity [23–25]. Barabasi [23] attributes this bursty non-Poisson character of human behavior to a priority-based queuing process. This view is supported by Jo et al. [24] who show that burstiness remains in mobile communication data even after circadian and weekly patterns have been removed, precluding the attribution of periods of inactivity to nights or weekends. They conclude that burstiness results from non-homogeneity in human task execution mechanisms. Kim et al. [26] conducted a study on a large dataset from North-American mobile phone users. The results suggest that the caller–callee behavior cannot solely be modeled using the Poisson distribution. Based on frequency of information exchange between the users, they classified the user-pairs into three categories characterized by the inter-arrival times between calls made between pairs. In a related study, Cardillo et al. [27] studied human proximity patterns in two data sets: the Reality Mining dataset and the co-location traces from INFOCOM'06. They found that proximity patterns from the MIT data contain both weekly and daily periodicity – most probably a result of how academic activities are scheduled at a university – while the INFOCOM'06 data showed only daily periodicity. Caridillo et al. extended this observation to study how cooperation emerges in a human society.

A patent from Google suggests that an adaptive contact list may detect contextual information for a given mobile phone user and may identify appropriate contact entries [28]. While studying the effects of two different UI adaptation techniques on user performance, Tsandilas and Schraefel [29] conclude that adaptation is always more effective, even when the accuracy of prediction is low. Bentley and Chen [30] found that the majority of contacts in a modern aggregated mobile phone book are rarely used. Their study shows that the five most frequently contacted alters represent 80% of phone and text communication. In addition, they found that a median of $Q = 60\%$ (six out of a total of $N = 10$ contacts) displayed in a "recent calls" list are amongst the most frequently contacted. While the authors use this latter statistic to argue against the efficacy of a "recent calls" list, it would be interesting to explore whether the value of Q increases for larger values of N, especially since the authors' results indicate an upward trend in Q as N increases from one to 10. Proponents of a "recent calls" list may argue that, in practice, these lists hold more than 10 entries. Based on their results, Bentley and Chen suggest a redesign of the content and representation of contact lists. A redesign of contacts book was proposed also in [31]. The data for Bentley and