Who's Calling? Demographics of Mobile Phone Use in Rwanda

Joshua E. Blumenstock
UC Berkeley
School of Information
jblumenstock@berkeley.edu

Dan Gillick
UC Berkeley
Computer Science Division
dgillick@cs.berkeley.edu

Nathan Eagle
Santa Fe Institute
nathan@santafe.edu

Abstract

We describe how new sources of data can be used to better understand the demographic structure of the population of Rwandan mobile phone users. After combining anonymous call data records with follow-up phone interviews, we detect significant differences in phone usage among different social and economic subgroups of the population. However, initial experiments suggest that predicting demographics from call usage, and vice-versa, is quite difficult.

Introduction

Despite the increasing ubiquity of mobile phones in the developing world, remarkably little is known about the structure and demographics of the mobile phone market. While a few qualitative studies have detailed social norms of phone use in specific communities (Donner 2007; Burrell 2009), and a handful of quantitative researchers have begun to analyze the dynamics of mobile phone networks in general (Onnela 2007; Eagle, Pentland, and Lazer 2009), data constraints have limited meaningful combination of the two.

Here we describe how electronic call data records (CDR) can be coupled with structured phone interviews to better understand the nature of mobile phone use in Rwanda. After introducing the data, we begin to investigate its structure through two parallel questions: First, what kinds of CDR features separate demographic categories? Second, can demographic features predict individual calling patterns?

Call Data Records:

Calling behavior	Men	Women	p-value
Total call duration	2.21	2.41	0.3498
Net call duration (out - in)	0.14	-0.19	0.0004
Int'l call duration	0.08	0.13	0.0586
Number of calls	4.83	4.66	0.5768
Net number of calls	0.23	-0.18	0.0045
Number of unique contacts	2.94	2.76	0.2979

Table 2: Differences in phone usage by gender. Daily averages reported; durations are in minutes.

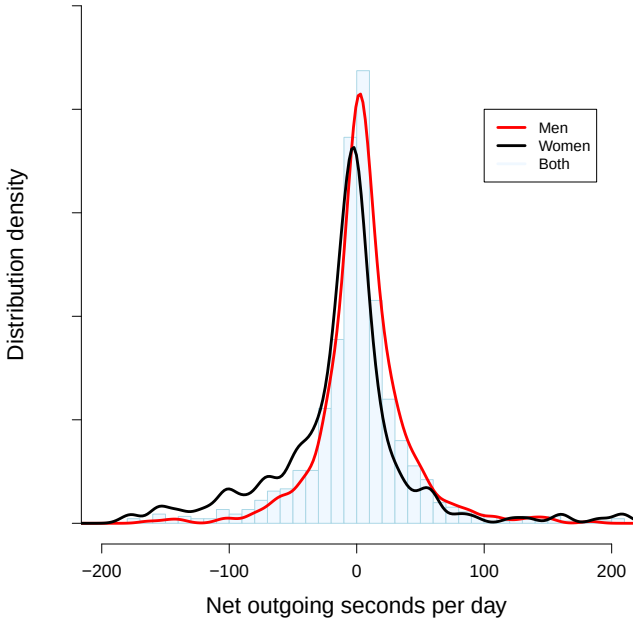

Figure 1: Net outgoing call time per day, by gender.

ences reported in Table 2, we thought it might be possible to infer gender (and other demographic attributes) based on CDR data alone. We trained a logistic regression model from 80% of the data and tested on the remaining 20%. Each experime3ealu-25065eporeadetobysplit
agis24322(v4(xpery43321ahht)-38526)7trfeariaes feade ver