

Dataset Name: Emergent smartphone users' dataset

Author Information

a. Principal Investigator Contact Information

Name: Aimal Rextin

Institution: COMSATS University Islamabad, Pakistan

Address: Department of Computer Science, COMSATS University, Islamabad, Pakistan

Email: aimal.rextin@comsats.edu.pk

b. Associate or Co-investigator Contact Information

Name: Shamailla Hayat

Institution: COMSATS University Islamabad, Pakistan

Address: Department of Computer Science, COMSATS University, Islamabad, Pakistan

Email: Shamaillahayat@upr.edu.pk, Shamaillahyt@yahoo.com

c. Alternate Contact Information

Name: Anas Bilal

Institution: Department of Computer Science, FAST National University of Computer and Emerging Sciences, Islamabad, Pakistan

Address: School of Computing FAST National University of Computer &

Emerging Sciences, A. K. Brohi Road, Sector H-11/4, 44000, Islamabad, Pakistan.

Email: anasbilal5773@gmail.com

Date of data collection (single date, range, approximate date)

The data was collected in December 2019 between 2019-12-01 - 2019-12-15

Geographic location of data collection

Dataset was collected from Islamabad and Kashmir

SHARING/ACCESS INFORMATION

1. Licenses/restrictions placed on the data: None

2. Links to publications that cite or use the data:

Paper will be published after the dataset is available publicly.

3. Links to other publicly accessible locations of the data: None

4. Links/relationships to ancillary data sets: None

5. Was data derived from another source? yes

If yes, list source(s): An aggregated version of data is provided here

to replicate the study findings, Actual datasets containing call and SMS logs of
of 30 Emergent and 30 Traditional users have not been shared due to privacy concerns
and policy of this repository. However, the original datasets could be provided by the
author on individual requests.

6. Recommended citation for this dataset: Citation of the paper: "Understanding the usability
issues in contact management of illiterate and semi-literate users"

DATA & FILE OVERVIEW

1. File List:

- I. The dataset contains three R scripts (plot_effectiveness_of_callogs.R, plot_efficiency_of_contact_search.R, plot_satisfaction_of_using_traditional_contactbook.R) to generate the plots of 'Usability Analysis'. The analysis is presented in section 'Usability of current contact-books and call logs by emergent users'.
- II. Two aggregated datasets named Emergent_User_Data.rar and Traditional_Users_Data.rar have been uploaded.

Emergent User Data.rar contains four .CSV files Named as:

1. Total_Contacts_and_Unintelligibly_Saved_Contacts_of_Emergent_Users
2. Characteristics_of_SMS_Data_of_Emergent_Users
3. Characteristics_of_Calling_Data_of_Emergent_Users
4. Percentage_of_Dialled_Contacts_through_Contactbook_of_Emergent_Users

Traditional User Data.rar contains four .CSV files Named as:

1. Total_Contacts_and_Unintelligibly_Saved_Contacts_of_Traditional_Users
2. Characteristics_of_SMS_Data_of_Traditional_Users
3. Characteristic_of_Calling_Data_of_Traditional_Users
4. Percentage_of_Dialled_contacts_through_Contactbook_of_Traditional_Users

These Two datasets contain the data needed to replicate the study findings presented in Section 'Objective data analysis of emergent and traditional users' of the paper.

2. Relationship between files, if important:

Files are independent of each other and each file can be used independently

3. Additional related data collected that was not included in the current data package: None
4. Are there multiple versions of the dataset? No

METHODOLOGICAL INFORMATION

1. Description of methods used for collection/generation of data:

This dataset containing the contact books and dual-channel (call, text) logs of 30 traditional and 30 emergent users, by using a customised Android application. The contact-book logs consisted of each contact's name, number, and total contacts of the user. While the call and text logs contained the following information for each communication event: contact name/number, time, date, duration, and status (incoming, outgoing, missed).

2. Methods for processing the data:

We used R as a programming tool to see how the composition of the contact-book of emergent users differ from those of traditional users in aspects like its size, prevalence use of special symbols, the proportion of dialled contacts through the phone-book, and percentage of unintelligible contact names, etc. All data analysis studies require data cleaning before analysing it to avoid misleading results. Hence, as a first step, the datasets were cleaned to maintain data integrity and then these aggregated datasets were computed by writing R scripts.

3. Instrument- or software-specific information needed to interpret the data:

These CSV files can be used in any analysis tool such as R, MATLAB etc., to replicate the study findings.

4. Standards and calibration information, if appropriate: None
5. Environmental/experimental conditions: None
6. Describe any quality-assurance procedures performed on the data: None, apart from necessary data cleaning process that have been described in the paper.
7. People involved with sample collection, processing, analysis and/or submission: First author of the paper.

DATA-SPECIFIC INFORMATION FOR: EmergentSmartphoneUsersDatasetReadMe

1. Number of variables:

- i. Data file "Total_Contacts_and_Unintelligibly_Saved_Contacts_of_Traditional_Users" in both datasets contains four variables.
- ii. Data file "Percentage_of_Dialled_contacts_through_Contactbook_of_Traditional_Users", in both datasets contains four variables.
- iii. Data file "Characteristic_of_Calling_Data_of_Traditional_Users", in both datasets contain six variables.

- iv. Data file " Characteristics_of_SMS_Data_of_Traditional_Users", in both datasets contain six variables.
2. **Number of cases/rows:** Each file has thirty (30) rows/records of thirty users.
3. **Variable List:**
 - i. Data file "Total Contaccs_and_Unintelligibly_Saved_Contacts_of_Traditional_Users" in both datasets contains the following four variables:
 - a. Users Ids (numeric value to indicate each user)
 - b. Total Saved Contacts (numeric value to indicate each users' total saved contacts in the phonebook)
 - c. Unintelligibly Saved Contacts (numeric value to indicate the total number of contacts that were stored using unintelligible contact names)
 - d. Percentage of Unintelligibly Saved Contacts (numeric value to indicate the percentage of unintelligibly saved contacts of each users)
 - ii. Data file Percentage_of_Dialled_contacts_through_Contactbook_of_Traditional_Users", in both datasets contains the following four variables:
 - a. Users Ids (numeric value to indicate each user)
 - b. totalOutgoingContacts (numeric value to indicate each users' total outgoing calls)
 - c. ContactsDialliedFromContactList (numeric value to indicate total outgoing calls of a users that were dialled through the phonebook)
 - d. PercentageOfDialledContactsFromContactlist (numeric value to indicate the percentage of total outgoing calls of a users that were dialled through the phonebook)
 - iii. Data file "Charateristic_of_Calling_Data_of_Traditional_Users", in both datasets contain the following six variables:
 1. Users Ids (numeric value to indicate each user)
 2. TotalCallsBeforeFilteration (numeric value to indicate each users' total calls before data cleaning process)
 3. TotalCallsAfterFilteration (numeric value to indicate each users' total ccalls after data cleaning)
 4. incomingCalls (numeric value to indicate each users' total incoming calls)
 5. outgoingCalls (numeric value to indicate each users' total outgoing calls)
 6. missedCalls (numeric value to indicate each users' total missed calls)
 - iv. Data file " Characteristics_of_SMS_Data_of_Traditional_Users", in both datasets contain the following six variables:
 - a. Users Ids (numeric value to indicate each user)
 - b. totalSms (numeric value to indicate each users' total SMS)
 - c. Incoming (numeric value to indicate each users' total incoming SMS)
 - d. Outgoing (numeric value to indicate each users' total outgoing SMS)
 - e. totalSolicitedSMS (numeric value to indicate each users' total solicited SMS i.e. SMS from valid contacts)
 - f. totalUnSolicitedSMS ()(numeric value to indicate each users' total unsolicited SMS, i.e. SMS from advertising companies etc.)
4. Missing data codes: None
5. Specialized formats or other abbreviations used: None

