Mobile Communication and Context Dataset

Mika Raento

Basic Research Unit, Helsinki Institute for Information Technology
Department of Computer Science, University of Helsinki
Mika.Raento@cs.Helsinki.FI

1 Dataset Contents

The Mobile Communications and Context Dataset contains logs of communication (calls and text messages) and context variables (location, profile and calendar) for a small number of people over a long period of time.

2 Sources

The data has been gathered for the Context project at Helsinki Institute for Information Science, funded by the Academy of Finland and is available through the project web site [1].

3 Data Characteristics

The data consists of an anonymized mobile phone communication log, location and profile changes as well as calendar events for a small number of people over periods varying from a couple of months to a year. Most of the data has been gathered with a software that runs in the background on commercially available mobile phone (the Nokia 7650), the calendar data being gathered separately from different calendar systems.

The communication log lists calls and text messages including date and time, duration, direction and originator/recipient of communication. Phone numbers and contact names have been anonymized: they have been replaced with sequentially assigned identifiers. The same identifier is always used for the same phone number/contact. The communication log data is not complete, but contains significant gaps.

The context data consists of location data in the form of GSM cell information (network operator, location-area-code (LAC) and cell identifier (CellId)). These have likewise been anonymized: values of each field has been individually replaced with sequential identifiers.

In addition to the values the logs contain stop and start markers. The software has not always been running and neither is the phone always on. In some cases there are no gaps in the data although there are start and stop markers. There is no totally reliable way of distinguishing between actual gaps and momentary

restarts, we have been using a rule of thumb that if the stop and start markers are more than ten minutes apart there is a real gap.

The calendar data lists the time, duration, type and description of events. The type and description have been anonymized. Calendar data is not available for every person.

4 Data Format

The communication and context logs are in separate files. The files are named `context-<id>.xml` and `comm-<id>.xml`, respectively. The `id` identifies the persons. XML [2] syntax is used for marking the logical structure in the files. Appendices B and C show example data.

The structure of both files is described by the Document Type Definition (DTD) in `log.dtd`, shown in Appendix A. The DTD includes comments for the different fields.

Datatypes for the fields are marked with `#FIXED` attributes on the corresponding elements. The datatype names and lexical representations are taken from XML Schema [3].

The calendar data is in file `cal-<id>.ics`. The standard iCalendar format [4] is used.

Although the XML syntax is verbose there are several reasons for using it:

- It includes a well-defined character set marking.
- Not all of our data fits in a flat row/column schema, with XML we can use a hierarchial structure for the values.
- Different kinds of data easily be included in the same file.
- With some care new kinds of data can be added to the structure without breaking existing applications.
- Parsers and other tools are widely available.
- It is human-readable.

5 Other Relevant Information

We plan to augment the dataset with data from more users and including more variables. Planned new context variables include application usage and general user activity on the mobile phone, and bluetooth device discovery.

If you are conducting research that requires access to the non-anonymized version of the data (or parts of it), please contact us and we can try to agree on a set of terms for such use.

We are planning to release the software used to gather the data. It is not ready for external release yet, but should you be interested you can contact us.

6 Past Usage

Kari Laasonen, Mika Raento and Hannu Toivonen: On-device Adaptive Location Recognition. To appear in the Proceedings of Pervasive 2004 [5].

References

1. Context project: Context recognition by user situation data analysis (Context). Website http://www.cs.helsinki.fi/group/context/ (2004)
2. W3C: Extensible Markup Language (XML) 1.0 (Third Edition). W3C Recommendation `http://www.w3.org/TR/2004/REC-xml-20040204/` (2004)
3. W3C: XML Schema Part 2: Datatypes. W3C Recommendation `http://www.w3.org/TR/2001/REC-xmlschema-2-20010502/` (2001)
4. Dawson, F., Stenerson, D.: RFC 2445: Internet calendaring and scheduling core object specification (iCalendar). IETF Proposed Standard (1998)
5. Laasonen, K., Raento, M., Toivonen, H.: Adaptive on-device location recognition. In: Proceedings of the 2nd International Conference on Pervasive Computing (Pervasive 2004). Lecture Notes in Computer Science, Springer-Verlag (2004) To appear.

A Document Type Definition

log.dtd

```
<!--
    log.dtd
    DTD for Mobile phone event data
    (c) 2004 Mika Raento
-->

<!ELEMENT events (event*)>

<!ELEMENT event (datetime, (location|profile|communication))>

<!ELEMENT datetime (#PCDATA)>
<!-- dt == datatype, XML Schema names and representations -->
<!ATTLIST datetime dt CDATA #FIXED "dateTime">

<!ELEMENT location (stop | (start, location.value?) | location.value) >
<!ELEMENT profile (stop | (start, profile.value?) | profile.value) >

<!ELEMENT communication ( stop | (start?, (comm.call|comm.sms), (comm.outgoing|comm.incoming),
    comm.duration, comm.number, comm.contact_name? ) ) >

<!ELEMENT location.value ( location.network, location.lac, location.cellid )>

<!ELEMENT profile.value ( profile.id )>

<!ELEMENT start EMPTY>
<!ELEMENT stop EMPTY>

<!-- GSM cell based location -->
<!ELEMENT location.network (#PCDATA) >
<!ATTLIST location.network dt CDATA #FIXED "string">
<!ELEMENT location.lac (#PCDATA)>
<!ATTLIST location.lac dt CDATA #FIXED "integer">
<!ELEMENT location.cellid (#PCDATA)>
<!ATTLIST location.cellid dt CDATA #FIXED "integer">

<!-- phone profile, 0-4 for Nokia 7650 -->
<!ELEMENT profile.id (#PCDATA)>
<!ATTLIST profile.id dt CDATA #FIXED "integer">

<!ELEMENT comm.call EMPTY>
<!ELEMENT comm.sms EMPTY>
<!ELEMENT comm.incoming EMPTY>
```

```
<!ELEMENT comm.outgoing EMPTY>
<!-- duration in secnds ifor calls -->
<!ELEMENT comm.duration (#PCDATA)>
<!ATTLIST comm.duration dt CDATA #FIXED "integer">

<!ELEMENT comm.number (#PCDATA)>
<!ATTLIST comm.number dt CDATA #FIXED "integer">
<!ELEMENT comm.contact_name (#PCDATA)>
<!ATTLIST comm.contact_name dt CDATA #FIXED "integer">
```
log.dtd

B Example Location and Profile Data

context-3.xml

```
<?xml version='1.0' encoding='iso-8859-1'?>
<!DOCTYPE events SYSTEM 'log.dtd'>
<events>
  <event>
    <datetime>20030227T144540</datetime>
    <location>
      <start />
      <location.value>
        <location.network>0</location.network>
        <location.lac>0</location.lac>
        <location.cellid>0</location.cellid>
      </location.value>
    </location>
  </event>
  <event>
    <datetime>20030227T144540</datetime>
    <profile>
      <start />
      <profile.value>
        <profile.id>1</profile.id>
      </profile.value>
    </profile>
  </event>
  <event>
    <datetime>20030227T144557</datetime>
    <profile>
      <profile.value>
        <profile.id>2</profile.id>
      </profile.value>
    </profile>
  </event>
  <event>
    <datetime>20030227T144634</datetime>
    <location>
      <stop />
    </location>
  </event>
  <event>
    <datetime>20030227T144634</datetime>
    <profile>
      <stop />
    </profile>
  </event>
</events>
```
context-3.xml

C Example Communication Log

comm-3.xml

```
<?xml version='1.0' encoding='iso-8859-1'?>
<!DOCTYPE events SYSTEM 'log.dtd'>
<events>
  <event>
    <datetime>20030203T093548</datetime>
    <communication>
      <start />
      <comm.call />
      <comm.incoming />
      <comm.duration>482</comm.duration>
      <comm.number>0</comm.number>
    </communication>
  </event>
  <event>
    <datetime>20030203T142753</datetime>
    <communication>
      <comm.sms />
      <comm.incoming />
      <comm.duration>0</comm.duration>
      <comm.number>1</comm.number>
      <comm.contact_name>0</comm.contact_name>
    </communication>
  </event>
  <event>
    <datetime>20030303T163312</datetime>
    <communication>
      <stop />
    </communication>
  </event>
</events>
```

comm-3.xml