Building a Personal Symbolic Space Model from GSM CellID Positioning Data

Filipe Meneses, Adriano Moreira

ubicomp@uminho
Mobile and Ubiquitous System Group
http://ubicomp.algoritmi.uminho.pt

Mobilware 2009 – Berlin, April 28th

Outline

- Motivation
- Positioning model
- Movement tracking
- Personal symbolic space model
- Experimentation

Motivation

- User’s position can be acquired using a number of different technologies:
  - GPS
  - Active Bat
  - Ubisense
  - WiFi networks

- Geometric positioning vs locations information
- The context of a person is more than a pair of coordinates or the name of a place. It is the type of place, with whom the person is, how often a person visits a place, etc.

Positioning Model (I)

- GSM
  - Mobile phone are well integrated in everyday life;
  - Turned on most of the time;
  - GSM networks are ubiquitous;

  - The mobile phone as a proxy to capture the context of a mobile user.

Positioning Model (II)

- GSM network
  - Telecommunications cellular network;
  - Made of cells with different size and shape;
  - The handset is linked to a cell – the active cell;
  - Each cell supports only a limited number of channels;
  - CellID positioning provides the handset location in a symbolic referential
    - Could be converted into the geographic position of the handset;
    - In fact, people on their living, deal with location in a symbolic referential (like “home”, “office”, etc.);

- The positioning model adopted in this work is based on the detection of movement within the space model of GSM cells.

Movement Tracking (I)

- Movement Tracking
  - Based on changes in the active cell;
  - When the terminal is stopped in a certain place, the sequence of active cells is limited to the set of cells that cover the terminal’s position;
  - Infer the user motion by analysing the changes in the serving cell and amount of time spent on each cell;

$$MobilityDistance(r_1, r_2) = \begin{cases} 0 & if \quad r_1 = r_2 \\ \dfrac{1}{time(r_1)} + \dfrac{1}{time(r_2)} & if \quad r_1 \neq r_2 \end{cases}$$

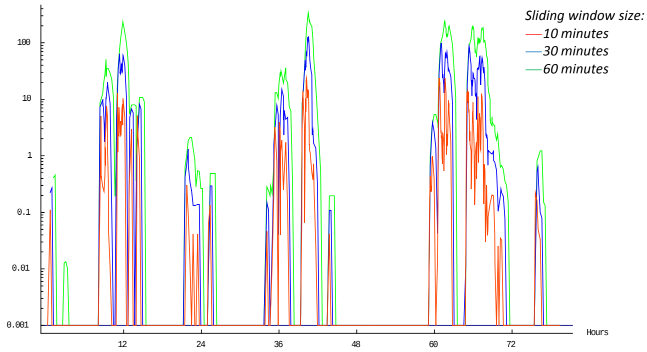
  - When the user is moving fast the time spent in a cell is small;

$$MobilityIndex(r_{1..n}) = \sum_{i=1}^{n} \sum_{j=1}^{i} MobilityDistance(\mathrm{r_i}, \mathrm{r_j})$$

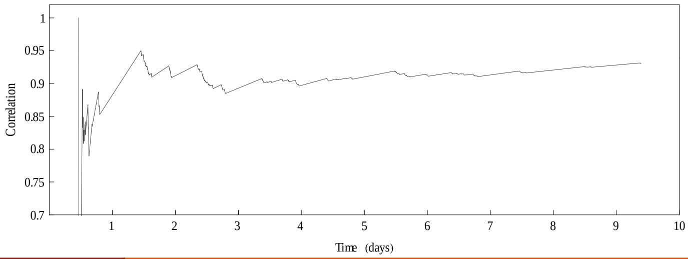
  - The user mobility level can be estimated by calculating the Mobility Index over a pre--defined period of time;
  - For a set of consecutive records it is possible to create a sliding window and calculate the Mobility Index as the time goes by;

Movement Tracking (II)

– Mobility index varies according to the size of the sliding window;
– To detect the user movements we defined a threshold based on sliding window size.



- Movement tracking validation:



Correlation between the mobility periods calculated from GPS data and from the Mobility Index

Sliding window: 10 minutes
Threshold: 6

Personal Symbolic Space Model (I)

- Mobility Index allows to detect when the user starts and stops moving;
  – Characterize the previously visited place: creating a **fingerprint** with the list of cells observed during the time the user was not moving

    $FP=\{\{\{cellID_1, timePercentage_1\},\{cellID_2, timePercentage_2\},\ldots,\{cellID_k, timePercentage_k\}\}, totaltime, timeStamp\}$

  – (However) Different visits to the same place create different fingerprints;

- Clustering similar fingerprints;
  – To create a cluster it is necessary to measure the similarity between fingerprints

$$FPDistance(FP^A, FP^B) = (0.5 \times HDistance(FP^A, FP^B)) + (0.5 \times SimilarityDistance(FP^A, FP^B))$$

Personal Symbolic Space Model (II)

- New clustering algorithm
  - Data do be clustered is symbolic (the fingerprints);
  - Clusters are to be discovered in real time;
  - The total number of clusters is not know in advance;
  - The number of clusters is the number of places visited by the user.

Personal Symbolic Space Model (III)

- The clustering process
  - 1^{st} fingerprint -> 1^{st} cluster
  - Calculate the similarity between the fingerprint and each cluster
    - Join the cluster if similarity is smaller than a pre-defined threshold
  - Create a new cluster
  - Replace an existent cluster

$$\mathrm{FCIR}(cl,\ fp) = \frac{KldgIdx(tt(fp))}{FgIdx(Age(cl,\ fp)) \times KldgIdx(tt(cl))}$$

- Familiarity index
  - Important places are usually visited more often and the user spends more time there (like home or office)

$$FmIdx(cl) = KldgIdx(cl) \times FgIdx(cl)$$

Experimentation (I)

- Data collection
  - Application on a Symbian phone;
  - Checks the cellID every eight seconds and records it on a log file;
  - Three different users collected data during several consecutive weeks;
  - Simultaneously manually recorded their movements on a diary;
- Clustering process
  - Similarity threshold: 65%

Experimentation (II)

- Trial users' results

	User A	User B	User C
Places visited by the user	4	27	29
Places detected	3 (16 clusters)	19 (21 clusters)	22 (30 clusters)
Places not detected	1	8	7
False positives	1 cluster	4 clusters	6 clusters

  - A detailed analysis of the achieved results can explain many of the errors

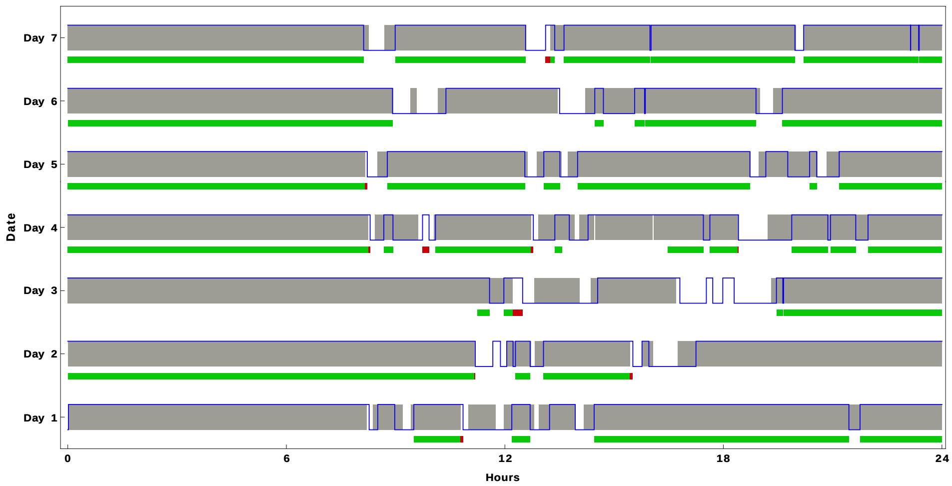
Experimentation (III)

Results achieved considering the visits made to each place

User A	2 places	53 visits	53 detected visits	100%
	1 place	6 visits	3 detected visits	50%
	1 place	3 visits	0 detected visits	0%
User B	18 places	58 visits	58 detected visits	100%
	3 places	71 visits	62 detected visits	87%
	6 places	9 visits	0 detected visits	0%
User C	12 places	18 visits	18 detected visits	100%
	10 places	133 visits	116 detected visits	87%
	7 places	11 visits	0 detected visits	0%

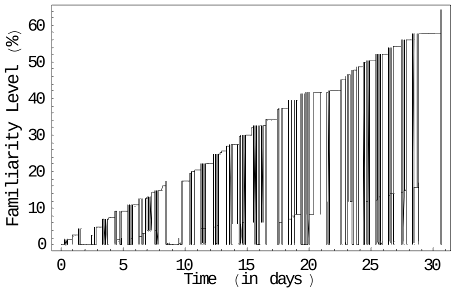
Experimentation (IV)

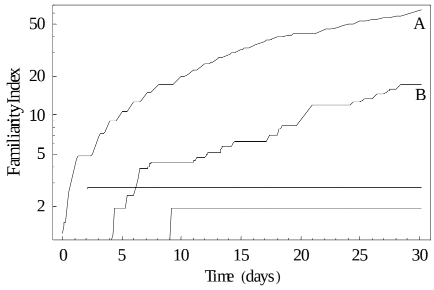
User agenda overlapped with the tracking and recognizer results



Experimentation (V)

- Using the Personal Symbolic Space Model
  - Identify the places visited by the user
  - A Familiarity Index can be valuable for different applications





Questions?

Thank you!