Proximity Classification for Mobile Devices Using Wi-Fi Environment Similarity

Alessandro Carlotto, Matteo Parodi, Carlo Bonamico, Fabio Lavagetto
DIST – University of Genova
alessandro.carlotto,matteo.parodi}@unige.it
{fabio,charlieb}@dist.unige.it

Massimo Valla
Telecom Italia Lab
massimo.valla@telecomitalia.it
Outline

- Objective
  - Compute distance by analysing signals from existing wireless networks.

- Approach

- Features Selection

- Gaussian Mixture Model Classifier

- Experimental tests and results
  - ‘Single scan’ approach.
  - ‘Best of three scans’ approach.

- Conclusions and future work
Objective

- Devise an algorithm to estimate device distance without any absolute information about locations.
  - Exploiting signals from existing Wi-Fi networks.
- Classify four levels of proximity
  - High - “same room”.
  - Medium - “nearby room / same floor”.
  - Low - “same building”.
  - No Proximity.
Approach

Key idea

- The closer two devices are, the more similar their radio environment will be.
- Estimate radio environment similarity by comparing a set of signal features.

1. Creation of a database of Wi-Fi scans.
2. Analysis of the collected data.
   - Choice of features useful to discriminate and simple to compute.
3. Choice of Gaussian Mixture Models (GMM) to represent different radio proximity levels.
4. Training of the models with a database of pairs of Wi-Fi scans collected in the corresponding situations.
Operating Scenario

- The algorithm is deployed on a remote Internet server.
- Each device sends MAC address, ID and signal strength of detected APs.
- The server receives scanned Wi-Fi data by devices.
- The server extracts required features from each pair of Wi-Fi data and selects the proximity category.
- The server sends back the proximity estimate.
Features (1)

- **Number of Common Access Points (NCAP)**
  - the number of APs in common between the two clients.

- **Number of Not-Common Access Points (NNAP)**
  - the number of not in common APs detected by the two clients.

- **Not-Common Access Points Difference (NAPD)**
  - it is the difference between the number of not in common access points detected by each of the two clients.

- **Mean Square Signal Strength Deviation (MSSSD)**
  - it represents the difference of signal strength on common access points measured from devices under consideration.

- **Wireless Common Power Percentage (WCPP):**
  - it’s the percentage of common radio power measured in dB by each device.
Features (2)

- NCAP = 3
  - AP1, AP3, AP4
- NNAP = 1 and NAPD = 1
  - Device A: None, Device B: AP2
- MSSSD = 152
  - Devices are at the same distance from each common AP.
- Wireless Common Power Percentage = 89%
Features (3)

- NCAP = 2
  - AP3, AP4
- NNAP = 2 and NAPD = 2
  - Device A: None, Device B: AP1, AP2
- MSSSD = 483
  - Devices are at different distance from each common AP.
- Wireless Common Power Percentage = 60%
Selection of a statistical model

- Gaussian Mixture Model.
- Density model with a number of K gaussian n-dimensional component functions.
- Each component is represented by the following adjustable parameters:
  - Weight.
  - Mean.
  - Covariance matrix.
Probability Densities of Wi-Fi environments’ GMMs.
Each GMM is composed by 3 bidimensional gaussian functions.
Classification algorithm

• When there are not AP in common between the two devices the classifier reports “No proximity”
Test setup

- Device used to test the system
  - Nokia N95
  - 802.11 b/g
- Test performed in a campus like environment

<table>
<thead>
<tr>
<th>Building characteristics</th>
<th>Building 1</th>
<th>Building 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number APs</td>
<td>24</td>
<td>16</td>
</tr>
<tr>
<td>Area</td>
<td>4050 m²</td>
<td>2400 m²</td>
</tr>
<tr>
<td>Avg. Area /APs</td>
<td>168.7 m²</td>
<td>150 m²</td>
</tr>
</tbody>
</table>
"Single Scan" Approach Results

- Correct level classification percentage: 68.7%

<table>
<thead>
<tr>
<th>Proximity Level</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>68.6%</td>
<td>25.7%</td>
<td>5.7%</td>
</tr>
<tr>
<td>Medium</td>
<td>14.8%</td>
<td>64%</td>
<td>21.2%</td>
</tr>
<tr>
<td>High</td>
<td>3.9%</td>
<td>21.9%</td>
<td>74.2%</td>
</tr>
</tbody>
</table>

Confusion Matrix
Improving the results

- We observed that the number of received APs strongly affects the classification performance.
- The phone sometime scans fewer Wi-Fi networks than those actually present, because of:
  - limited sensitivity of the smartphone’s Wi-Fi hardware
  - long beacon interval of some APs.
“Best of Three Scans” Approach

Results

- Groups of three scans collected in the same place and within thirty seconds.
- Algorithm fed with the scan that has the highest number of APs detected.
- Correct level classification percentage: 88.2%

Confusion Matrix

<table>
<thead>
<tr>
<th>Proximity Level</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>96.5%</td>
<td>2.9%</td>
<td>0.6%</td>
</tr>
<tr>
<td>Medium</td>
<td>13.6%</td>
<td>77.8%</td>
<td>8.6%</td>
</tr>
<tr>
<td>High</td>
<td>1.7%</td>
<td>8.1%</td>
<td>90.2%</td>
</tr>
</tbody>
</table>

Percentage of correct decision as function of the total number of APs.
Example Application

- Proximity aware buddy list for Symbian smartphones.
- The application uses the proximity classifier system to detect nearby users.

- The building’s icon on the left of the names indicates that the technology used to obtain proximity is Wi-Fi.
- The black bars on the right represent the three proximity levels.
Conclusions

- The classification system discriminates among three different proximity levels between two devices.
  - Classifier chooses the correct level using simple features extracted by devices’ Wi-Fi AP scans.
- One-time training is required to create the statistical classification models.
  - The training is not environment-dependent.
- The algorithm reaches a percentage of correct classification near to 90%, achieving the best results where at least 3-4 APs are detected.
Future Work

- The algorithm will be extended to classify proximity using information derived from other technologies.
  - Bluetooth, GSM Cell-Ids, etc.

- The application will be extended including user proximity alerts and integration with APIs offered by popular social networks.
  - LinkedIn, Facebook, etc.