Classification of movement data and user’s activity recognition via mobile phones

Spyridoula Tragopoulou
Harokopio University of Athens
Magdalini Eirinaki
San Jose State University

Iraklis Varlamis
Harokopio University of Athens
varlamis@hua.gr
Contents

• The idea
• Related applications
• Architecture
• Implementation
• Experiments
• Conclusions and future work
THE IDEA

Motivation

Big Data from Cheap Phones

Collecting and analyzing information from simple cell phones can provide surprising insights into how people move about and behave—and even help us understand the spread of diseases.
THE IDEA

The challenge

- Collect user activity information with a smart-phone
  - Position, speed, altitude and time information
- Analyze collected data using the smart-phone
  - Real-time classification of user’s movement
- Visualize user trail
- Store user information in a data repository for future usage
  - Extract habits and make recommendations

Android Application

http://galaxy.hua.gr/~it20934/
COMPARISON

Related Applications

**MyTracks**

**RunKeeper**
COMPARISON

Features

<table>
<thead>
<tr>
<th>Application</th>
<th>Motion Classification</th>
<th>Show on map</th>
<th>Biosignals</th>
</tr>
</thead>
<tbody>
<tr>
<td>MyTracks</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>RunKeeper</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>GPSTracker</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
APPLICATION ARCHITECTURE

GPSTracker Architecture
IMPLEMENTATION

Data Recording

- Android Service running in background
- Features
  - Longitude, Latitude, Average speed, Smoothed average speed, Near Metro Station, Altitude, Altitude change, Timestamp, Time zone, Day of the week, GPS Signal status
- An instance \((movement)\) every 30 sec
Movement Classification

- Recognize the type of each user movement
- Movement Types: Walking, Running, Biking, Driving, Metro, Bus, Motionless

- Collect training data for every type
- Build a classification model
- Store the model in the device
- Classify every new movement instance
IMPLEMENTATION

Training Data Collection

![Maps and applications showing GPS data and move type selection.](image)
IMPLEMENTATION

Classification algorithm

- Weka API for Android
  https://github.com/rjmarsan/Weka-for-Android
- Tree-based classifiers
  - Fast predictions on mobile devices
  - Light model
  - Good performance
- RandomForests had the best performance
IMPLEMENTATION

Real-time Trail Visualization

- Google Maps API for Android
IMPLEMENTATION

Visualization of stored trails

Select File to Open

Thu Sep 19 12:19:43 EEST 2013.arff
Thu Sep 19 13:16:03 EEST 2013.arff
Mon Sep 16 18:14:03 EEST 2013.arff
Mon Sep 16 19:45:59 EEST 2013.arff
Mon Sep 16 21:08:32 EEST 2013.arff
Mon Sep 16 21:11:23 EEST
IMPLEMENTATION

Repository of user trails

- Upload user files to an online data repository
- Use Dropbox API so that data are stored in a private repository for each user
EXPERIMENTS

Training Dataset

- 4518 training samples
- 10-fold cross validation on the training data
- Tree-based classifiers
- RandomForests had the best performance (92.81±0.99 at 99% confidence level)
- Confident model for movement predictions

<table>
<thead>
<tr>
<th>Movement Type</th>
<th>Number of training samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking</td>
<td>770</td>
</tr>
<tr>
<td>Running</td>
<td>177</td>
</tr>
<tr>
<td>Biking</td>
<td>343</td>
</tr>
<tr>
<td>Driving</td>
<td>650</td>
</tr>
<tr>
<td>Metro</td>
<td>1256</td>
</tr>
<tr>
<td>Bus</td>
<td>534</td>
</tr>
<tr>
<td>Motionless</td>
<td>788</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>J48</td>
<td>90.73</td>
</tr>
<tr>
<td>LMT</td>
<td>85.79</td>
</tr>
<tr>
<td>RandomForests</td>
<td>92.81</td>
</tr>
<tr>
<td>REPTree</td>
<td>87.61</td>
</tr>
<tr>
<td>RandomTree</td>
<td>91.05</td>
</tr>
</tbody>
</table>
EXPERIMENTS

Learning Curves - Random Forests

20% test : 904 instances
80% training
CONCLUSIONS

Future Work

• Extension of the set of features for motion classification
• Long term analysis of user information
• Post-processing of movement data from multiple users
  ▫ using a shared data repository (and users’ consent)
  ▫ extract significant places and user habits
• Geospatial extension of SQLite RDBMS including public transportation stops and routes, parks, malls and shopping areas
• Personalized notification system
Thank you for your attention?

Questions?
### Accuracy on test data

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy %</th>
<th>Accuracy % (no speed smoothing)</th>
</tr>
</thead>
<tbody>
<tr>
<td>J48</td>
<td>89.12</td>
<td>88.43</td>
</tr>
<tr>
<td>LMT</td>
<td>80.47</td>
<td>84.33</td>
</tr>
<tr>
<td>RandomForest</td>
<td>92.14</td>
<td>92.00</td>
</tr>
<tr>
<td>REPTree</td>
<td>86.55</td>
<td>86.71</td>
</tr>
</tbody>
</table>
## Confusion matrix on test data

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>g</th>
<th>&lt;classified as</th>
</tr>
</thead>
<tbody>
<tr>
<td>135</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>11</td>
<td>a = Walking</td>
</tr>
<tr>
<td>1</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>b = Running</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>70</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>c = Biking</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>127</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td>d = Driving</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>243</td>
<td>2</td>
<td>1</td>
<td></td>
<td>e = Metro</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>93</td>
<td>1</td>
<td></td>
<td>f = Bus</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>139</td>
<td></td>
<td>g = Motionless</td>
</tr>
</tbody>
</table>