Sensor-Based Single-User Activity Recognition

Blaž Kažič, Jan Rupnik

AI Lab
Jozef Stefan Institute
Ljubljana, Slovenia
Motivation

• Improving Smartphone Technology

• Better Build-In Sensors

• Motion Co-Processor
Possible Applications

- Healthcare
- Business
- Transportation, ...
Outline – Workflow

- Collecting Data
- Feature Extraction
- Classifier Selection
- Application Results
Collecting Data

- Mobile App
  - SensorLogger
- Extracting Data
  - 30Hz
- Sample
  - 10s
- Training Data Set
  - 1750 samples (5h)

<table>
<thead>
<tr>
<th>Person</th>
<th>Standing</th>
<th>Walking</th>
<th>Running</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>80</td>
<td>77</td>
<td>80</td>
</tr>
<tr>
<td>b</td>
<td>79</td>
<td>116</td>
<td>94</td>
</tr>
<tr>
<td>c</td>
<td>79</td>
<td>78</td>
<td>59</td>
</tr>
<tr>
<td>d</td>
<td>92</td>
<td>86</td>
<td>68</td>
</tr>
<tr>
<td>e</td>
<td>68</td>
<td>92</td>
<td>86</td>
</tr>
<tr>
<td>f</td>
<td>166</td>
<td>79</td>
<td>78</td>
</tr>
<tr>
<td>g</td>
<td>68</td>
<td>63</td>
<td>62</td>
</tr>
<tr>
<td>Sum</td>
<td>632</td>
<td>591</td>
<td>527</td>
</tr>
</tbody>
</table>

Table: Training data set
Orientation Problem

Sitting - Person a

Sitting - Person b

Sitting - Person c

Amplitude vs. Time (s) for different axes and persons.
Raw Accelerometer Data

Person a - Sitting

Person a - Walking

Person a - Running

Person a - All activities

0.5s

0.3s

Amplitude

Time (s)

Amplitude

Time (s)

Amplitude

Time (s)

Amplitude

Time (s)

AI Lab, JSI

7.10.2013
Extracting Features

• All Features Set
  • \{mean, standard deviation, variance, median, root mean square, skewness, kurtosis, 25th percentile, 75th percentile, inter-quartile, mean crossing rate, dominant frequency, DFTs energy, spectral entropy, xy correlations, xz correlation, yz correlation\}

• Simplified Features Set
  • \{mean, standard deviation, 75th percentile, dominant frequency, xy correlation\}

• Mean & StdDev Set
  • \{mean, standard deviation\}
Mean & StdDev
Dominant Frequency

Walking Dominant Frequencies

Aver Dom Freq = 2 Hz

Running Dominant Frequencies

Aver Dom Freq = 2.9 Hz
Classifier Selection

- **Weka Toolkit**
  - Open Source Machine learning Software
  - Collection of machine learning algorithms for data mining tasks
  - ARFF (Attribute Relationship File Format)

- **Tested Machine Learning algorithms**
  - Decision Tree (J.48)
  - Naive Bayes (NB)
  - K-Nearest Neighbor (IBK)
  - Support Vector Machine (SMO)
  - Neural Network (Multilayer Perceptron)
Classifier Evaluation

Feature and Classifier Comparison

<table>
<thead>
<tr>
<th>Classifier</th>
<th>All features</th>
<th>Simplified features</th>
<th>Mean &amp; StdDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Tree (J.48)</td>
<td>87.3%</td>
<td>88.2%</td>
<td>82.7%</td>
</tr>
<tr>
<td>Naive Bayes (NB)</td>
<td>89.2%</td>
<td>93.2%</td>
<td>80.9%</td>
</tr>
<tr>
<td>k-Nearest Neighbor (IBK)</td>
<td>71.9%</td>
<td>71.6%</td>
<td>86.7%</td>
</tr>
<tr>
<td>Support Vector Machine (SMO)</td>
<td>82.1%</td>
<td>85.9%</td>
<td>85.9%</td>
</tr>
<tr>
<td>Neural Network (Multilayer Perceptron)</td>
<td>73.3%</td>
<td>82%</td>
<td>88.6%</td>
</tr>
</tbody>
</table>
Naive Bayes Classifier

- **Experiment**
  - Learning set: 1750 Labeled Samples
  - Input: 294 New Labeled Samples
  - Result: 283 Correctly Classified (96%)

<table>
<thead>
<tr>
<th>Labeled input data</th>
<th>Recognized activity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Running</td>
</tr>
<tr>
<td>Running</td>
<td>92</td>
</tr>
<tr>
<td>Walking</td>
<td>0</td>
</tr>
<tr>
<td>Sitting</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2: Confusion matrix
Practical Application

- Activity Diary
- Twelve hours of data recorded
- Filtering data
Results

Activity Diary - Continuous

- Running
- Walking
- Sitting

Hour

08 09 11 12 13 14 15 16 17 18 19 20
Results

Activity Diary - Continuous

Activity Diary - Filtered
Future Work

• Increase number of recognizable activities
  • Sitting, walking, running, standing, hiking, cycling
  • driving, shopping...

• Include more sensors
  • GPS, gyroscope, compass, ...

• Larger training set, larger testing set

• Sampling frequency, window sizes
Thank You for Your Attention