A Notebook Sensory Data Set for Context Recognition

Harald Radi, Rene Mayrhofer, Alois Ferscha

Institut für Pervasive Computing, Johannes Kepler University Linz
Altenberger Str. 69, A-4040 Linz, Austria
{mayrhofer,radi,ferscha}@soft.uni-linz.ac.at

Abstract. For a qualitative and quantitative assessment of context prediction and recognition methods, real-world data sets are inevitable. By collecting sensor data on a single notebook over a period of a few months we got a rather large log file of homogeneous and heterogeneous features reflecting the users activities during this time frame. In this paper we present which devices were exploited as sensors, which information was logged and how this information was stored for further processing by classification algorithms.

1 Introduction

Beside artificially generated test cases we also need some real-world data sets for a precise analysis of the meaning of available sensor informations. Thus, the first step is to log every available information that has a potential impact on the context to be recognized. By logging even redundant information we ensure that no information is lost during the first step of data gathering and thereby eliminate any failure cause before classification.

The data set was generated using the means of our context recognition and prediction framework proposed in [1]. The framework was deployed as a service on a notebook running Windows XP, thus logging was active whenever the notebook was turned on. The notebook was used for daily work on a regular basis during the past four months, sampling data every 30 seconds.

The resulting log file is a ASCII file with roughly 90000 samples and a small XML file containing meta data to describe the different states of heterogeneous sensors that are referenced in the log file by unique ids. A possible meta data set could look like Fig. 1 while the corresponding log file entry would be the id of the element.

2 Sensors and Features

The configuration shown in Fig. 2 was used to initialize the framework and perform the logging. Features are specified by a unique id containing the name of the feature provider which corresponds to a sensor, a dot separator and the name of the feature that should be logged. A short description of the features meaning follows.

```
<feature id="Wlan.ActiveEssid">
    <element id="0"><![CDATA[]]></element>
    <element id="1"><![CDATA[nme]]></element>
    <element id="2"><![CDATA[universe]]></element>
    <element id="3"><![CDATA[ipi]]></element>
</feature>
```

Fig. 1. Meta data

```
<features>
    <feature id="Time.Timestamp" />
    <feature id="ActiveWindow.ActiveWindow" />
    <feature id="Audio.Mean" />
    <feature id="Audio.Peaks" />
    <feature id="Audio.Band.0" />
    <feature id="Audio.Band.1" />
    <feature id="Audio.Band.2" />
    <feature id="Audio.Band.3" />
    <feature id="Audio.Band.4" />
    <feature id="Audio.Band.5" />
    <feature id="Audio.Band.6" />
    <feature id="Audio.Band.7" />
    <feature id="Audio.Band.8" />
    <feature id="Audio.Band.9" />
    <feature id="Audio.Band.10" />
    <feature id="Audio.Band.11" />
    <feature id="Audio.Band.12" />
    <feature id="Audio.Band.13" />
    <feature id="Audio.Band.14" />
    <feature id="Audio.Band.15" />
    <feature id="Power.Plugged" />
    <feature id="Wlan.ActiveEssid" />
    <feature id="Wlan.ActiveMode" />
    <feature id="Wlan.ActiveSignalLevel" />
    <feature id="Wlan.ActiveMacAddress" />
    <feature id="Wlan.Peers" />
    <feature id="Wlan.NumPeers" />
    <feature id="GSM.CellID" />
</features>
```

Fig. 2. Configuration

- *Timestamp* Timestamp indicating when the log entry occurred. It can be used to calculate the month, day of week, hour or any other information that can be derived from a time stamp.
- *ActiveWindow* The title of the application that had the focus.
- *Mean* The average signal level measured on the laptops build-in microphone.
- *Peaks* The number of peaks that have been detected in the microphones input signal.
- *Band.** The level of the corresponding frequency band after performing a FFT. The total number of available frequency bands was set to 16 prior to sampling.
- *Plugged* A binary feature signaling whether the Laptop was plugged or battery driven.
- *ActiveEssid* Name of the connected WLAN Network at the time of sampling.
- *ActiveMode* Signals whether the WLAN card was operating in ad-hoc or infrastructure mode. This is especially interesting as Windows automatically switches the WLAN card into ad-hoc mode if no access point can be reached.
- *ActiveSignalLevel* Signal strength of the signal available to the WLAN card.
- *ActiveMacAddress* The MAC address of the access point that was used.
- *Peers* A list of MAC addresses containing all WLAN peers that were reachable during the scan.
- *NumPeers* The length of the Peers list.
- *CellID* The GSM cell id the cell phone was assigned to while being connected to the notebook via Bluetooth.

3 Conclusions

The data set was gathered on a notebook that was used for daily work and should therefore reflect different situations in a students day like "being at the university", "at work" or "at home"; or even more specifically "surfing in the Internet", "working on the context recognition project", etc. The data set contains sensor information over a period of four months. The laptop was usually switched on regardless whether it was used or not, thus also periods of inactivity should be recognizable.

Further analysis of the data set will contain a statistical examination of the different features to find out which features are meaningful for the description of a context and which are ambiguous and can be left out. Weighting the features before classification could probably also lead to more meaningful clusters, this has to be inspected.

References

1. Mayrhofer, R., Radi, H., Ferscha, A.: Feature extraction in wireless personal and local area networks. In Agha, K.A., Omidyar, C.G., eds.: The Proceedings of The Fifth IFIP-TC6 International Conference on Mobile and Wireless Communications Networks (MWCN 2003), World Scientific (2003) 195–198