

Problem Sheet 4

Submission Deadline: 04.02.2008

1 Problem

When dealing with machine learning several terminologies, methods, and concepts are widely used. Define and explain the following terminologies and their applications:

- a) Supervised Learning vs. Unsupervised Learning
- b) Training Set vs. Test Set
- c) Feature
- d) Confusion Matrix

Note: Use the material provided on the website as a starting point to answer the questions.

2 Points

2 Problem

Recognizing context information, e.g. environmental context, is very important in wearable computing. Figure 1 shows a diagram of the components involved in a typical pattern recognition system that can be used to recognize context.

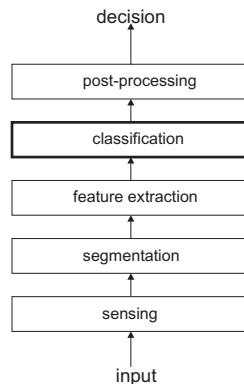


Figure 1: Components of a pattern recognition system

As Figure 1 shows there are five different components involved in a recognition process. The recognition process starts with input sensing by gathering sensor data and ends by post processing and decision making upon the results of a classifier.

In this problem we focus on practicing basic signal analysis and segmentation for hand gesture recognition. The next problem 2 will then concentrate on the classification process of those gestures.

For recording *training* and *test sets* we performed different hand gestures with the right hand while having attached an acceleration sensor to the back of the hand. There were four different gestures performed: **"Thumb up"**, **"Thumb left"**, **"Thumb right"**, and **"something else"**. While performing the gestures we always started from the "Thumb up" position and then turned the hand either left or right and back again to the "Thumb up" position.

For recording the *training set* we performed 20 "left" gestures by starting each from "Thumb up" position followed by 20 "right" gestures. The speed in which the gestures where performed was not fixed. The recorded data is available for download on the lecture homepage.

For recording the *test set* we randomly performed 40 gestures ("left" or "right"). The speed in which the gestures were performed was not fixed. The recorded data is also available for download as well as a general description of the structure of the recorded data sets.

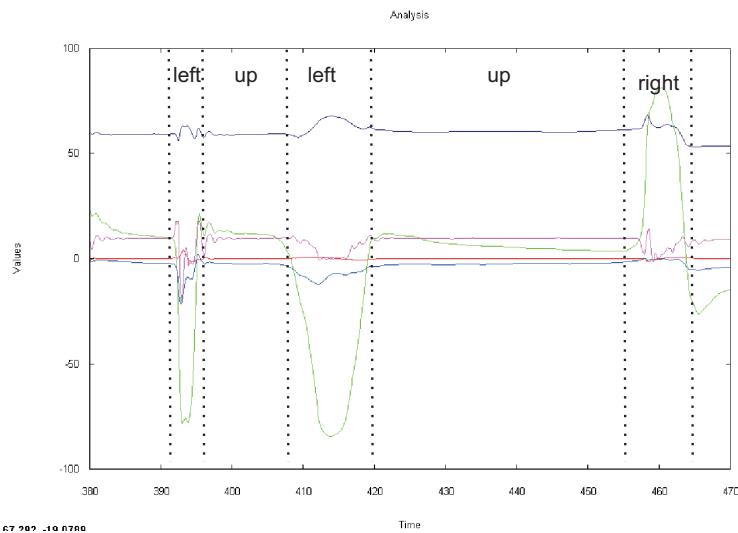


Figure 2: Segmented and annotated data

Figure 2 shows an example chart of manually segmented and annotated raw sensor data gathered from the XSens MTx¹ acceleration sensor while performing the different gestures.

- a) Start analyzing the raw sensor data given in the training set by plotting an appropriate subset of the measurements and decide which measurement type characterizes the gestures best. While doing so, use for annotation the four different hand gesture names "thumb left", "thumb right", "thumb up", and "something else". Prepare the chart similar to the example given in figure 2. What are the problems when segmenting the data visually and by hand respectively?
- b) After that, classify the training set manually by adding a new column to the training data file where you map each row to one of the four gestures, e.g. by introducing different single letters for each gesture. For this task you basically have to extend your segmentation subset from your chart to the entire training set.

3 Points

¹Further information can be found at <http://www.xsens.com> or the provided technical documentation

3 Problem

The WEKA² machine learning tool is a free open source collection of machine learning algorithms for data mining tasks.

- a) Use your manually classified training set from the previous problem including the features already existing in the test set to build a WEKA compatible ARFF training set file.
- b) Explain how a decision tree classification algorithm works in principle.
- c) Train a decision tree classifier (e.g. WEKA J.48) with your training set and discuss the results given by the confusion matrix. What recognition rates have you achieved on the training set?
- d) Use your already trained decision tree classifier and let it classify the test set provided (make sure that you have converted it into an ARFF format). Discuss the results of the classification process and write down the decision tree used by your classifier.
- e) Describe and explain which features were used by the classifier. What can you derive from this?

5 Points

Note: The submission of problem sheets has to be done by e-mail to hwitt@tzi.de with the following subject:

[Lecture Wearable Computing] Problem Sheet No.{number}. If problems require the implementation of software it has to be submitted along with the problem sheet containing documented source code, compiled binaries, and 2 small scripts (for Windows and Linux) to run the application.

²<http://www.cs.waikato.ac.nz/~ml/weka/>