Title: Using quantitative measurements from mobile phone for predicting user activity

Introduction:

The goal of activity recognition is to identify the actions carried out by a person given a set of observations of itself and the surrounding environment. Smart phones can be used to acquire sensor data to enable physical activity recognition and number of applications have been studied by researchers in activity recognition. [1]

Data mining or knowledge discovery is the process of discovering interesting regularities in large masses of data. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. [2]

Exploratory analysis and data mining techniques were used in order to predict subject activity based on quantitative measurement from mobile phone. Our analysis suggests that the measurements can definitely be used for classifying subject activity. The most challenging classifications tasks are related to differentiating between different types of walking activity and, on the other hand, between sitting and standing.

Methods:

For our analysis we used the data on 30 volunteers between ages 19-48 years. The data were downloaded from Coursera Data Analysis course web page [3] on February 27, 2013 using the R programming language [5]. The loaded data was slightly processed to make them easier to load into R. The original raw data is available in Machine Learning Repository. [4]

The data contained 10299 measurements of the subjects performing six activities wearing a smartphone on the waist: walking, walking upstairs, walking downstairs, sitting, standing and laying. There were 561 different attributes, which included 3-axial linear acceleration and 3-axial angular velocity. The obtained dataset was partitioned into two sets, where data from subjects 1, 3, 5, and 6 were selected for generating the training data and data from subjects 27, 28, 29, and 30 for generating the test data. The training data consisted of 1315 measurements and the test data consisted of 1485 measurements. No data transformations were needed for the data.

Tables and plots of the observed data were examined in order to perform exploratory analysis. Support Vector Machine (SVM) method was used for classification and regression tasks. [6]

Results:

The results of the classification using SVM model are reported in Figure 1.

Activity	Laying	Sitting	Standing	Walk	Walk down	Walk up	*Sensitivity %*
Laying	**292**	1	0	0	0	0	*99.7*
Sitting	0	**183**	3	0	0	0	*98.4*
Standing	1	80	**280**	0	0	0	*76.3*
Walk	0	0	0	**185**	0	0	*100.0*
Walk down	0	0	0	19	**190**	8	*87.6*
Walk up	0	0	0	25	10	**208**	*85.6*
Precision %	*99.7*	*69.3*	*99.0*	*80.8*	*95.0%*	*96.3*	

Figure 1. Confusion table of the classification results for the test data. Differentiating between different types of walking activity and between sitting and standing are most difficult classification tasks, resulting lower sensitivity rates.

Conclusions:

Our analysis suggests that the measurements acquired by a smart phone can be used for classifying subject activity. However, differentiating between different types of walking activity and between sitting and standing are most difficult classification tasks. Using larger training data set could provide more accurate model for better prediction of those activities. In addition, further tuning of the SVM parameters and combining models from alternative classification methods, such as classification trees, could result better classification accuracy rates.

References

1. Wikipedia "Activity Recognition" Page. URL: http://en.wikipedia.org/wiki/Activity_recognition . Accessed 2/27/2013.
2. Wikipedia "Data Mining" Page: URL http://en.wikipedia.org/wiki/Data_mining . Accessed 2/27/2013
3. Coursera Data Analysis. Samsung phone data. URL: https://spark-public.s3.amazonaws.com/dataanalysis/samsungData.rda. Accessed 2/27/13
4. Machine Learning Repository. URL: http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones . Accessed 2/27/2013.
5. R Core Team (2012). "R: A language and environment for statistical computing." URL: http://www.R-project.org

6. S. R. Gunn. Support vector machines for classification and regression.Tech. rep., University of Southampton, UK, 1998.