Recognition of human activity based on smartphone sensor data.

Jakob L.K. Gerstenlauer & Yunxuan Dou   Universidad Politècnica de Catalunya, 06.06.2017

1 Abstract

Smartphones have become an ubiquitous phenomenon. While carrying them with us, they record our movements and these measurements can later be used to infer the activity of the owner, which has a range of possible applications. We used a publicly available data set with pre-processed sensor data and labels of human activity. We analyzed the data set using multivariate methods to, first, visualize and describe the input data and, second, build a predictive model of human activity. We evaluated the prediction model based on an independent test set and found an overall accuracy of 83.2% (92.6% if standing and sitting are considered as one category).

2 Motivation

Smartphones have become an ubiquitous phenomenon. While carrying them with us, they record our movements using internal accelerometer and gyroscope sensors. These measurements can be used to infer the activity of the owner of the smartphone which has a range of possible applications such as e.g. detecting accidents of elderly persons or monitoring the activity of users in special health apps [Ang+13b]. Smartphone apps might also use sensor data to classify the activity of users in order to adapt their behaviour accordingly. They might e.g. increase the resolution of the screen to facilitate the readability of text when the user is walking. It might also make sense to adapt the frequency of displaying commercial adds to user activity. In the combination with other pieces of information such as the geographical location of the smartphone additional insights are possible. A navigation app might e.g. be able to warn a user when he is going downstairs instead of upstairs inside a building. With this enumeration of possible applications, we have indicated the relevance of statistical models predicting user activity from smartphone sensor data a task that has already been approached in a number of studies[Ang+13b; Rey+13; All+06; MS10; Kha+10]. In this work, we used a publicly available data set with preprocessed smartphone sensor data that was collected during a experiment where different persons were instructed to perform six different types of activities, such as walking, sitting, standing and lying[Ang+13a]. We analyzed this data set with the objective of, first, reducing the dimensionality and finding latent variables and, second, predicting the human activity using a random forest classifier. Our ultimate goal was to assess the feasibility of predicting human activity based on smartphone sensor data.

3 Methods and Results

3.1 Data Set and Primary Data Processing

The data was retrieved from the “Human Activity Recognition Using Smartphones Dataset”[Ang+13a]. The different data sets and a detailed description of the experiment are available at the UCI machine learning data base [Jor13] and on a GitHub account which also contains the *R* code[Jak17]. The data had already been separated into a training (N=7352, p=561, s=21) and a test set (N=2947, p=561, s=9), randomly assigning subjects to either the training or the test set. As expected, the data sets contained the same number of predictors p, but differed in the number of observations N and the number of subjects s. The observations had been hand labeled during data collection. Each subject performed six activities while wearing a smartphone on the waist (Table 1). The distribution of these activity labels was not equally distributed (Table 1).

Label:	1	2	3	4	5	6
Data Set	WALKING	UPSTAIRS	DOWNSTAIRS	SITTING	STANDING	LAYING
Train	1226	1073	986	1286	1374	1407
Test	496	471	420	491	532	537

Table 1: Frequency table of activity classes in the training and the test set. Here UPSTAIRS and DOWNSTAIRS refer to walking up stairs or walking down stairs.

The processing of the raw sensor data is described in detail in the data set documentation: “The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low frequency components, therefore a filter with 0.3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain. See 'features_info.txt' for more details. ” Here, the important insight is that observations from the same subject and the same label represent a time series of overlapping measurement windows. We, thus, have to expect auto-correlations between these observations.

3.2 Secondary Data Processing

In this section, we describe data processing steps conducted by ourselves. In the training set, we tested for a significant effect of all predictors in a one way anova model with the activity class label as response. We found six out of 561 predictors without a significant effect and excluded these variables from all further analysis. Then we used the *lofactor* function from the *DMwR* R package to calculate an outlier score and exclude the top four outliers. Then, we calculated weights for all observations in the test set in order to give equal importance to all subject $\times$ class crossings. We used these weights as metric in the principal component analysis.

3.3 Principal Component Analysis and Feature Extraction

We performed a principal component analysis based on the centered and scaled data matrix. We decided to retain 59 principal components based on the 90% cumulative explained variance rule (based on the Kaiser rule we would have retained 60 components). Based on these 59 selected principal components, we performed the following feature selection schedule in parallel for the training and the test data set (the test data was first transformed in the same way as the training data using the stored vectors of means and standard deviations): For each principal component, we selected all variables with a loading higher than 0.8 or lower than -0.8. If no such variables were detected we decreased the threshold to 0.7. Then we computed a new aggregate feature by averaging all selected variables (variables with negative loadings were first multiplied with -1). Fifteen principal components without associated variables showing a loading of at least 0.7 or -0.7 were ignored. In parallel, the 44 new features were also calculated for the test set – after being centered and standardized according to the means and standard deviations of the training set.

3.4 Cluster Analysis

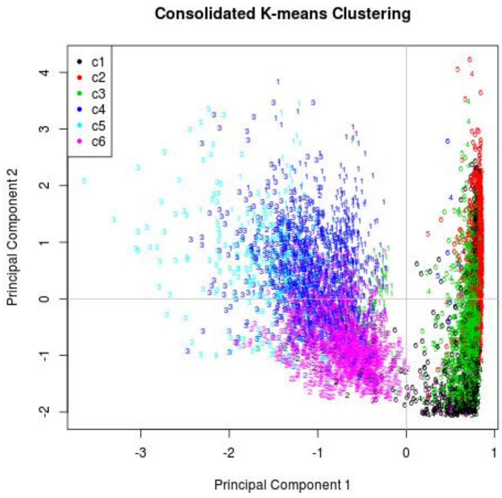
Based on the 44 extracted features, we performed a hierarchical cluster analysis with k=6 because we had six classes. We first transformed the new features into an euclidean distance matrix and then used the Ward aggregation method to perform a hierarchical cluster analysis using the *hclust* function with *method* argument *ward.D2*. We further consolidated the partition by using the final centroids of the six clusters as starting point for k-means using the *kmeans* function. The explained variance in terms of class labels of these consolidated clusters was only 20.0%. The results in terms of class membership are depicted in Table 2. Clusters 1,2,3 mostly contain observations with labels 1,2,3 (walking) and clusters 4,5,6 mostly contain observations with labels 4,5,6 indicating stationary activities. There is, however, no clear connection between clusters and specific labels, we can only make out some tendencies.

empirical class	1	2	3	4	5	6
cluster						
1	0	20	1	271	120	545
2	0	2	1	618	549	694
3	72	15	10	376	690	161
4	871	95	496	2	5	3
5	100	44	203	0	0	0
6	183	897	274	19	8	3

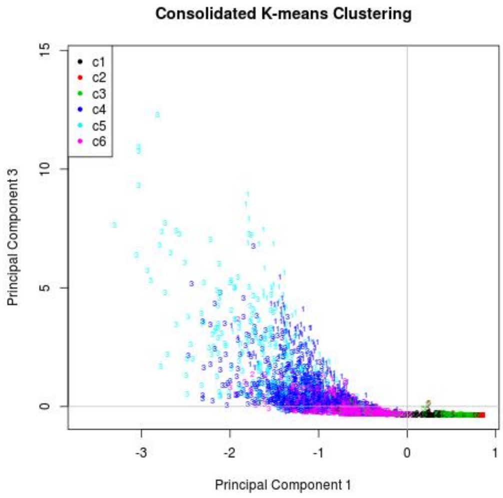
Table 2: Table of class membership in clusters by k-means with centroids by hierarchical clustering.

3.5 Bagging the Cluster Analysis

Based on the poor performance of the original cluster analysis and based on the fact that it is not possible to directly include weights in the standard cluster analysis functions, we tried to improve the initial cluster analysing using a weighted bootstrapping schedule: We repeatedly (N=100) sampled from the training



(a) First factorial plane.



(b) Second factorial plane.

Figure 1: Projection of observations into the first and second factorial planes. Consolidated k-means clusters are indicated with different colours. Note the clear separation between clusters 1,2,3 and 4,5,6 by the first principal component.

data set with replacement drawing the same number of observation N as in the original data set. Each observation was sampled with probability equal to its weight. Then we rerun the original two-stage cluster analysis as described above for each probabilistic bootstrap sample. We calculated the centroids of the six clusters as the median over all bootstrap samples. In order to assign observations from the training set to the clusters, we calculated the cosine similarity between the individuals and the cluster centroids. Then we assigned individuals to the cluster with maximum cosine similarity. The results in terms of class membership are depicted in Table 3. Lamentably, the bagging approach did not work as intended: In comparison to Table 2, clusters 1 and 6 are not able to clearly separate the blocks of classes 1,2,3 (walking) and 4,5,6 (stationary). However, we used the bootstrapping approach to assess the variability of the explained variance of hierarchical cluster analysis (mean=17.71%, SD=0.34%) and consolidated k-means (mean=19.73%, SD=0.31%).

empirical class	1	2	3	4	5	6
cluster						
1	6	0	9	620	584	666
2	14	0	2	61	147	18
3	60	1	10	121	207	50
4	1126	600	936	5	6	0
5	2	1	0	242	275	241
6	18	471	28	237	153	431

Table 3: Class membership in clusters of training data with centroids determined by median of centroids of all bootstrap samples.

3.6 Projecting the Test Individuals into Clusters

We assigned observations from the test set to the clusters based on the maximum cosine similarity between the centroids determined in the regular consolidated cluster analysis and the 44 new extracted features in the test set. The results in terms of class membership are depicted in Table 3. As for the training set, clusters clearly separated observations from the two blocks of activity classes describing movement or stationary activities, they were, however, not able to separate between the different walking or stationary activities.

empirical class	1	2	3	4	5	6
cluster						
1	0	24	0	99	85	185
2	1	1	4	237	171	325
3	15	4	8	151	275	23
4	362	33	196	0	0	3
5	64	48	143	0	0	0
6	54	361	69	4	1	1

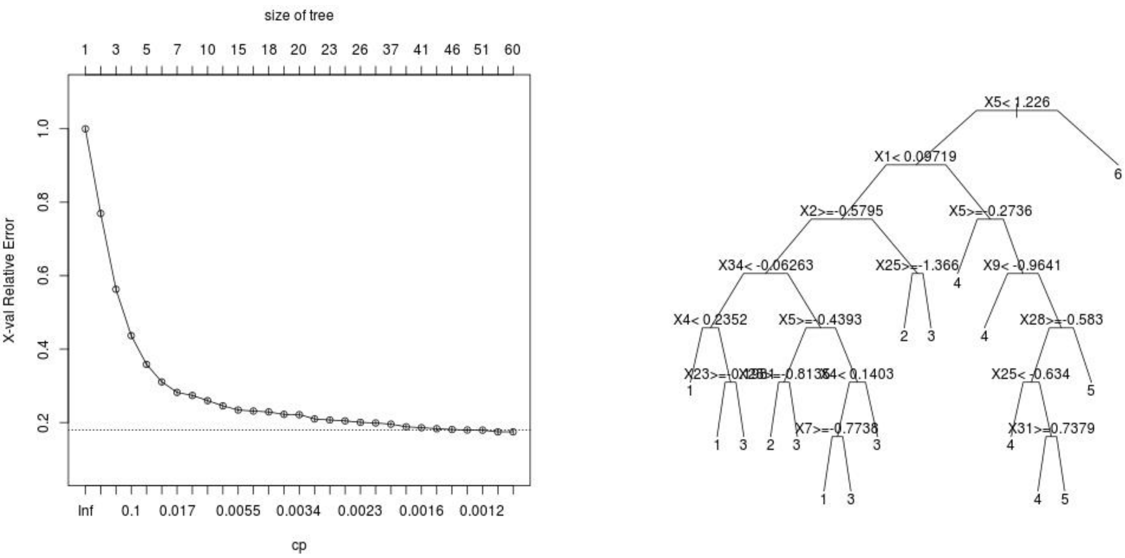
Table 4: Class membership in clusters of test data with centroids determined by centroids of consolidated k-means cluster analysis.

3.7 Regression Tree Analysis

In order to prepare the prediction model based on random forests, we used the *rpart* function from the *rpart* library to fit a binary regression tree to the training data. Based on ten-fold cross-validation, we decided to set the complexity parameter *cp* to 0.005 (Figure 2a). The corresponding pruned regression tree contained 17 terminal nodes (Figure 2b). This optimal number of nodes gave us an indication of a sensible number of leave nodes for the random forest analysis.

3.8 Random Forest Analysis

Regression and classification trees are high-variance, low-bias models for which bagging works especially well[HTF09, p. 587]. That is why a bagged version of trees, called random forest, was proposed that bases its prediction on an ensemble of trees[Bre01]. We used the *randomForest* function from the *randomForest* package to fit a random forest with 500 trees to the 44 new features extracted from the training data. We restricted the number of nodes of individual trees to a maximum of 15 trees and we restricted individual nodes to a minimum size of 10 observations. We used the *classwt* argument to enforce equal weights for all six classes. Furthermore, we set the *mtry* argument to 10, i.e. at each split the trees would randomly chose ten out of the 44 features and split the data based on the feature maximally reducing node impurity. The resulting ensemble of classification trees showed an impressive accuracy for the test set (Table 5). The accuracy was especially high for the activity "laying", where only one observation was misclassified. Although the accuracy for the other activities was not as high, the results are nevertheless satisfactory because misclassifications mostly occurred to very similar activities: No observation



(a) Cross-validated prediction error for different cp values.

(b) Pruned regression tree.

Figure 2: Regression tree analysis.

of "walking" behaviour (classes 1,2,3) was assigned to a class describing stationary behaviour (classes 4,5,6). The two classes describing sitting (class 3) and standing (class 4) were difficult to separate, which makes sense and does not indicate a weakness of the classifier. Only nine variables showed a variable importance (percentage increase in misclassification rate) which was higher than 2%. Not surprisingly, the five most important inputs corresponded to the first five principal components.

prediction	1	2	3	4	5	6	Error rate
Empirical class							
1	457	13	26	0	0	0	0.079
2	41	412	18	0	0	0	0.125
3	62	48	310	0	0	0	0.262
4	0	2	0	276	213	0	0.438
5	0	7	0	63	462	0	0.132
6	0	1	0	0	0	536	0.002

Table 5: Random forest confusion matrix for the test set. The global accuracy is 83.2%. If we do not differentiate between standing and sitting, the accuracy increases to 92.6%.

4 Discussion

4.1 Interpreting the Latent Constructs

The principal component analysis revealed 59 significant components explaining 90% of the inertia. This indicates that the 561 original features are indeed representing a high dimensional space which can not be reduced below a dimensionality of 60 without loosing considerable information. In order to improve the ease of interpretation and to further reduce the dimensionality of the data set, we applied a tailor-made feature extraction method: We filtered the significant principal components for components with a high loading (equivalent to correlation in this case) of original variables. Then we replaced the principal component by the mean of all the original variables associated to the component. Based on this approach, we were able to further reduce the dimensionality to 44. These 44 features were used for all further analyses and they can be interpreted as a common factor represented by the associated variables. Based on the fact that it is impossible for us to interpret all 59 significant principal components or the 44 extracted features, we will focus on the first principal component which explained 50.96% of total inertia (second principal component: 6.68%). As can be seen in Figure 1, the first component is able to partially separate observations of the different classes. Surprisingly, 240 out of 555 original variables (some variables had been excluded because there was no effect) showed an absolute value of the loading greater 0.8. This shows that these 240 variables are all very highly correlated, and that the first principal component efficiently represents the variation of these 240 variables. Based on the sheer number, we are not able to go into the details of the semantics of these variables here. However, the first principal component probably describes the intensity of both horizontal and vertical movements which are associated with short accelerations and deceleration events. This would explain why the first component clearly separates classes associated with walking from classes associated with stationary activities but is unable to distinguish between the different types of walking or stationary activity.

4.2 Interpreting the Clusters

The six clusters can not be unequivocally linked to each of the six activity classes (Table 2). However, we can clearly associate cluster 1, 2, 3 to the block of activities associated with walking (classes 1, 2, 3) and the clusters 4, 5, 6 to the block of activities associated with stationary behaviour (classes 4, 5, 6). As we learned from the confusion matrix of the random forest model (Table 4), classes 4 (sitting) and 5 (standing) are particularly hard to differentiate probably because the subjects are in a vertical position and they are not moving. We also learned from this confusion table that the three moving activities (classes 1,2,3) are difficult to differentiate among each other. This finding explains why we can not find clear associations between clusters and activity classes. Nevertheless there are some tendencies, cluster 4 is linked to class 1 and cluster 6 is linked to class 2.

4.3 Caveats

Our data analysis is based on data from only 30 subjects. Observations from the same subject are not independent. Instead they represent a time-series of sensor data from the same device attached to the same subject. Both devices and human individuals may display singular behaviour. The findings of our work are limited, given that the classifier was only trained with data from 21 human individuals and evaluated on test data from 9 human individuals. Throughout or analysis, we did not take the auto-correlation of data into account. We only weighted observations from activity classes and subjects equally in the principal component analysis. We also tried to take these weights into account when running the cluster analysis and implemented a weighted bootstrapping sampling scheme. This adapted bagging approach to our cluster analysis did however not perform very well and we ignored the results from this analysis. Based on the low generalization errors of the final random forest classifier, we are nevertheless confident that it was not inappropriate to ignore the auto-correlation structure of the data. Future studies should be based on more representative data sets with data from more subjects and devices.

4.4 Conclusions

In this work, we were able to show that it is feasible to predict the activity of smartphone users based on a restricted set of 44 features. However, the accuracy of the final model was conditional on the activity class: The final model was very accurate in distinguishing walking from stationary behaviour. The model was, however, less accurate in predicting if a subject was walking in the plane, walking downstairs, or walking upstairs. For stationary postures, the classifier was close to perfect when it came to predicting if the subject was lying. It was, however, more difficult to distinguish between sitting and standing. Our conclusion is, thus, that sensor data can be efficiently used to distinguish between three types of activities: lying, a stationary vertical position (standing or sitting), and walking. Predicting if a subject is walking upstairs or downstairs is possible but error-prone. Based on these results, a number of applications could use our classifier to predict user activity and adapt accordingly.

5 Acknowledgments

We appreciate the suggestions of Tomàs Aluja regarding principal components based feature extraction.

References

Articels

[Bre01] Leo Breiman. “Random forests”. In: *Machine learning* 45.1 (2001), pp. 5–32.

[All+06] Felicity R Allen et al. “Classification of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models”. In: *Physiological measurement* 27.10 (2006), p. 935.

[Kha+10] Adil Mehmood Khan et al. “A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer”. In: *IEEE transactions on information technology in biomedicine* 14.5 (2010), pp. 1166–1172.

[MS10] Andrea Mannini and Angelo Maria Sabatini. “Machine learning methods for classifying human physical activity from on-body accelerometers”. In: *Sensors* 10.2 (2010), pp. 1154–1175.

[Ang+13a] Davide Anguita et al. “A Public Domain Dataset for Human Activity Recognition using Smartphones.” In: (2013).

[Ang+13b] Davide Anguita et al. “Energy Efficient Smartphone-Based Activity Recognition using Fixed-Point Arithmetic.” In: *J. UCS* 19.9 (2013), pp. 1295–1314.

[Rey+13] Jorge Luis Reyes-Ortiz et al. “Human Activity and Motion Disorder Recognition: towards smarter Interactive Cognitive Environments.” In: (2013).

Books

[HTF09] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2009.

Online

[Jor13] Jorge L. Reyes-Ortiz, Alessandro Ghio, Luca Oneto, Davide Anguita and Xavier Parra. *Human Activity Recognition Using Smartphones Dataset*. 2013. URL: http://archive.ics.uci.edu/ml/machine-learning-databases/00240 (visited on 06/02/2017).

[Jak17] Jakob Gerstenlauer. *SmartphoneMobility*. 2017. URL: https://github.com/jakobgerstenlauer/SmartphoneMobility (visited on 06/04/2017).