FitActivity Codebook

Rohith Shankar

5/9/2020

1. Merges the training and the test sets to create one dataset

Read the features file and assign proper column names
features <- read.table("UCI HAR Dataset/features.txt", sep=" ", col.names = c("x","functions_used"))
(561 rows and 2 columns)

Read the activity file and assign proper column names
activity_labels <- read.table("UCI HAR Dataset/activity_labels.txt", sep=" ", col.names = c("code","activity_label"))
(6 rows and 2 columns)

Read the train subject file and assign proper column names
subject_train <- read.table("UCI HAR Dataset/train/subject_train.txt", col.names = "subject")
(7352 rows and 1 columns)

Read the test subject file and assign proper column names subject_test <- read.table("UCI HAR Dataset/test/subject_test.txt", col.names = "subject")
(2947 rows and 1 columns)

Read the y train file and assign proper column names
y_train <- read.table("UCI HAR Dataset/train/y_train.txt", col.names ="code")
(7352 rows and 1 columns)

Read the x train file and assign proper column names x_train <- read.table("UCI HAR Dataset/train/x_train.txt", col.names=features$functions_used)
(7352 rows and 561 columns)

Read the x test file and assign proper column names x_test <- read.table("UCI HAR Dataset/test/x_test.txt", col.names=features$functions_used)
(2947 rows and 561 columns)

Read the y test file and assign proper column names y_test <- read.table("UCI HAR Dataset/test/y_test.txt", col.names ="code")
(2947 rows and 1 columns)

Union the train and text x files
x <- rbind(x_train, x_test);
(10299 rows and 561 columns)

Union the train and text y files
y <- rbind(y_train, y_test);
(10299 rows and 1 columns)

Union the train and test subject files
subject <- rbind(subject_train, subject_test)
(10299 rows and 1 columns)

Join the x,y and subject files
mergedData <-cbind(subject,x,y);
(10299 rows and 563 columns)

2. Extract only measurements on the mean and standard deviation for each measurement

Extract only those columns that have are mean or standard deviantions
meanSubset1<-select(mergedData,subject, code, contains("mean"), contains("std"));

Join mean/std data with the activity subset
mergedData1<-select(merge(meanSubset1, activity_labels, by.x="code", by.y="code"),-code)
(10299 rows and 88 columns)

Review the names of the metric columns
names(mergedData1)

Repalce the cryptic variable names with descripitive names
names(mergedData1)<-gsub("Acc", "Accelerometer", names(mergedData1))
names(mergedData1)<-gsub("Gyro", "Gyroscope", names(mergedData1))
names(mergedData1)<-gsub("BodyBody", "Body", names(mergedData1))
names(mergedData1)<-gsub("Mag", "Magnitude", names(mergedData1))
names(mergedData1)<-gsub("^t", "Time", names(mergedData1))
names(mergedData1)<-gsub("-freq()", "Frequency", names(mergedData1))
names(mergedData1)<-gsub("angle", "Angle", names(mergedData1))
names(mergedData1)<-gsub("gravity", "Gravity", names(mergedData1))
names(mergedData1)<-gsub("^f", "Freuqency", names(mergedData1))
names(mergedData1)<-gsub("^tBody", "TimeBody", names(mergedData1))
names(mergedData1)<-gsub("-mean()", "Mean", names(mergedData1))
names(mergedData1)<-gsub("-std", "STD", names(mergedData1))

3.Creates a second, independent tidy data set with the average of each variable for each activity and each subject

summarise_all(group_by(mergedData1, subject, activity_label), funs(mean))
(180 rows and 88 columns)

Write the new tidy dataset to a file
write.table(mergedData1, "SummarizedActivityData.txt")